Functional constraints on SoxE proteins in neural crest development: the importance of differential expression for evolution of protein activity

Authors: Eric M. Lee[♥], Tian Yuan[♥], Reyna D. Ballim[♣], Kristy Nguyen[♥], Robert N. Kelsh[♣], Daniel M. Medeiros[♦], and David W. McCauley[♥♠]

[♥]Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Norman, OK
[♦]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO
[♣]Department of Biology and Biochemistry, University of Bath, UK

[♠]Author for correspondence: dwmccauley@ou.edu

Keywords: lamprey, zebrafish, SoxE, gene duplication, neural crest, functional divergence

Abstract

Vertebrate SoxE genes (*Sox8*, *9*, and *10*) are key regulators of neural crest cell (NCC) development. These genes arose by duplication from a single SoxE gene in the vertebrate ancestor. Although SoxE paralogs are coexpressed early in NCC development, later, *Sox9* is restricted to skeletogenic lineages in the head, and *Sox10* to non-skeletogenic NCC in the trunk and head. When this subfunctionalization evolved and its possible role in the evolution of the neural crest are unknown. Sea lampreys are basal vertebrates that also possess three SoxE genes, while only a single SoxE is present in the cephalochordate amphioxus. In order to address the functional divergence of SoxE genes, and to determine if differences in their biochemical functions may be linked to changes in neural crest developmental potential, we examined the ability of lamprey and amphioxus SoxE genes to regulate differentiation of NCC derivatives in zebrafish *colourless* (*cls*) mutants lacking expression of *sox10*. Our findings suggest that the proto-vertebrate *SoxE* gene possessed both melanogenic and neurogenic capabilities prior to SoxE gene duplication. Following the agnathan-gnathostome split, lamprey *SoxE1* and *SoxE3* largely lost their melanogenic and/or enteric neurogenic properties, while gnathostome SoxE paralogs have retained functional conservation. We posit that this difference in protein subfunctionalization is a direct consequence of the independent regulation of SoxE paralog expression between the two lineages. Specifically, we propose that the overlapping expression of gnathostome SoxE paralogs in early neural crest largely constrained the function of gnathostome SoxE proteins. In contrast, the largely non-overlapping expression of lamprey SoxE paralogs allowed them to specialize with regard to their DNA-binding and/or protein interaction properties. Restriction of developmental potential among cranial and trunk neural crest in lampreys may be related to constraints on SoxE activity among duplicates, but such

specialization does not appear to have occurred in gnathostomes. This highlights an important difference in the evolution of SoxE activity between these two divergent vertebrate lineages and provides insights for understanding how cell fate restriction in different NCC populations may be dependent on subfunctionalization among SoxE duplicates.

Introduction

Neural crest cells (NCC) are migratory, pluripotent cells that give rise to a wide range of vertebrate-specific cell types (Hall, 1999a; LeDouarin and Kalchiem, 1999). The emergence of NCCs during vertebrate evolution is thought to be linked to the evolution of the vertebrate craniofacial skeleton which facilitated the transition from filter feeding to a predatory lifestyle (Gans and Northcutt, 1983). Consistent with this view, invertebrate chordates are filter feeders and lack bona fide NCCs (Holland and Short, 2008; Medeiros, 2013; Putnam et al., 2008).

Lampreys occupy a basal position in vertebrate phylogeny; they are primitively jawless (agnathan) vertebrates that diverged from the lineage leading to gnathostome (jawed) vertebrates over 450 million years ago (Janvier, 1996; Kuratani et al., 2002). Like other vertebrates, lampreys possess multi-gene families that play critical roles in NCC development, whereas many of these exist as single copy genes in invertebrates (Meulemans and Bronner-Fraser, 2007; Yu et al., 2008).

The Sry-related HMG box (Sox) family of transcription factors is found throughout the animal kingdom (Wegner, 1999). Sox genes are related by the presence of a conserved high mobility group (HMG) box DNA-binding domain (Laudet et al., 1993; Prior and Walter, 1996). Sox proteins regulate gene expression by binding the consensus sequence $^{A}/_{T}\,^{A}/_{T}CAA^{A}/_{T}G$ (Mertin et al., 1999; van de Wetering et al., 1993), and also through their interactions with other

protein partners (Kondoh and Kamachi, 2010). Chordates share a highly conserved set of Sox proteins belonging to the SoxA, SoxB, SoxC, SoxD, SoxE, SoxF, and SoxH subfamilies, with multiple members of subgroups (B – F) often present in vertebrates (Bowles et al., 2000). For example, vertebrates possess three members of the SoxE subfamily (*Sox8*, *Sox9*, and *Sox10*), and at least two of these (*Sox9* and *Sox10*) are required for NCC specification and differentiation (Carney et al., 2006; Dutton et al., 2001; Guth and Wegner, 2008; Hong and Saint-Jeannet, 2005; Spokony et al., 2002; Wegner and Stolt, 2005), while less is understood about possible roles for *Sox8* in neural crest.

Sox9 regulates expression of the *Col2a1* gene encoding the major extracellular matrix protein, type-II collagen, in gnathostome vertebrate cartilage (Akiyama et al., 2002; Bell et al., 1997; Zhang et al., 2006). *SOX9* defects in humans lead to Campomelic dysplasia characterized by major defects in cartilage and bone (Wagner et al., 1994). Two *Sox9* paralogs (*sox9a* and *sox9b*) exist in zebrafish due to teleost-specific gene duplication (Cresko et al., 2003; Postlethwait et al., 2004; Yan et al., 2002; Yan et al., 2005). Together, *sox9a* and *sox9b* perform the functions of the single ancestral *Sox9* (Chiang et al., 2001; Li et al., 2002; Rau et al., 2006; Yan et al., 2005). The zebrafish s*ox9a* loss of function mutant *jellyfish* (*jef*) exhibits major loss of craniofacial cartilage elements, while loss of both *sox9a* and *sox9b* leads to more severe cartilage defects, suggesting partitioning of roles (i.e., subfunctionalization) between the two genes (Lynch and Force, 2000; Yan et al., 2002; Yan et al., 2005).

Sox10, required for multiple stages of NCC development, is first expressed during specification of NCCs in the dorsal region of the neural tube (Cheng et al., 2000; Honore et al., 2003). It persists in migrating NCCs, later becoming restricted to the peripheral nervous system (PNS), and melanocyte lineages (Aoki et al., 2003; Bondurand et al., 1998; Bondurand et al.,

2000; Britsch et al., 2001; Dutton et al., 2001; Southard-Smith et al., 1998). Mutations in human *SOX10* lead to variants of Waardenburg syndrome and Hirschsprung's disease, characterized by defects in pigmentation and the enteric nervous system (ENS), as well as dysmyelination syndromes (Pingault et al., 1998). The zebrafish *sox10* loss-of-function mutant *colourless* (*cls*) lacks differentiated pigment cells, and enteric ganglia, and has reduced numbers of dorsal root ganglia (DRG) (Carney et al., 2006; Dutton et al., 2001; Elworthy et al., 2005; Kelsh and Eisen, 2000; Kelsh and Raible, 2002) .

Early in development, *Sox8*, *Sox9*, and *Sox10* are co-expressed and in some models (e.g., frogs) may be functionally redundant during early neural crest specification. For example, in the chick, *Sox9* and *Sox10* are expressed in the anterior open neural plate at HH stage 10, and by stage 12, all three SoxE paralogs are expressed in premigratory NCCs along the dorsal midline (Cheung and Briscoe, 2003). In the 14 hr. post fertilization (hpf) zebrafish embryo, *sox9b* is expressed in premigratory NCCs (Yan et al., 2002), while by 16 hpf, *sox10* is also expressed in cranial and trunk premigratory NCCs (Dutton et al., 2001). Interestingly, zebrafish *sox9a* is not expressed in the neural crest (Yan et al., 2005). In *Xenopus*, both *Sox9* and *Sox10* are expressed in developing neural crest along the anteroposterior axis (Aoki et al., 2003; Spokony et al., 2002). However, the degree to which SoxE proteins have diverged functionally with respect to differences in their ability to regulate gene activity within different NCC lineages later in development is not clear (Cossais et al., 2010a; Kellerer et al., 2006; O'Donnell et al., 2006; Taylor and Labonne, 2005). In *Xenopus,* both *Sox9* and *Sox10* can increase the formation of NCCs, and both induced the formation of melanoblasts (Taylor and Labonne, 2005). Evidence from mice suggests *Sox8* is able to replace some glial and neurogenic (sensory and sympathetic) neural crest derivatives, but *Sox8* was unable to rescue melanogenesis in *Sox10*-deficient mice

(Kellerer et al., 2006). Such results suggest SoxE genes retain partially overlapping functions, but functional differences in protein activity have also arisen following SoxE gene duplication. Furthermore, the degree to which neofunctionalization and/or subfunctionalization of duplicated SoxE paralogs was a driving force in the diversification of NCC derivatives during early vertebrate evolution remains unclear. In this paper, we have focused on neural crest-specific activities dependent on divergent roles of SoxE proteins, and how functional diversification may be related to changes in developmental expression.

In the sea lamprey (*Petromyzon marinus*), SoxE genes are expressed in partially overlapping domains during neural crest formation (Tahara Stage 23). *PmSoxE1* is expressed in premigratory cranial NCCs while at the same stage, *PmSoxE2* (orthologous to *Sox10*; see results below) is expressed along the entire anteroposterior axis as well as in early migrating NCCs (Sauka-Spengler et al., 2007). *PmSoxE3*, the lamprey *Sox9* ortholog, is expressed in only a few premigratory cranial NCCs at stage 23 (Lakiza et al., 2011). Previously, we showed differential SoxE protein function during sea lamprey development. The loss of *PmSoxE2* resulted in a complete lack of pigment cells following antisense morpholino mediated knockdown. However, knockdown of *PmSoxE1* did not have the same phenotypic effect on melanogenesis (Lakiza et al., 2011).

Comparison of *SoxE* genes between jawed and jawless vertebrates reveals differences in *SoxE* expression patterns between these two groups during neural crest development. Functional studies have suggested overlapping activity of SoxE proteins in jawed vertebrates, while differences in protein activity are seen in lampreys. Taken together, these observations suggest the possibility that functional differences in the neural crest roles of *SoxE* genes may be related to evolutionary changes in their expression during neural crest development. A prediction of this

hypothesis is that SoxE genes restricted in their neural crest expression may be unable to compensate for loss of function among paralogs, while SoxE genes with overlapping neural crest expression domains may retain redundant functional activity in regulating cell fate specification. In order to test this hypothesis, we examined the functional activity of SoxE genes present in the sea lamprey by expressing constructs in a zebrafish *cls* mutant background. Presently, expression studies in lamprey to examine the evolution of SoxE gene functions remain technically challenging due to the long development time in the lamprey (12 days to melanogenesis), and the unavailability of genetic null backgrounds. However, the availability of a zebrafish mutant lacking *sox10* expression (Dutton et al., 2001), their rapid development time, ease of manipulation, and optical clarity for analysis provide a strong rationale for expression of foreign constructs from lamprey, as well as other *SoxE* genes from amphioxus and frog (Van Otterloo et al., 2012). Here, we show that foreign *SoxE* DNA constructs expressed in zebrafish are functional but are only partially able to rescue loss-of-function phenotypes in the *sox10* (*cls*) mutant background. Our results suggest that activities of the ancestral *SoxE* gene were likely co-opted by neural crest cells early in vertebrate evolution. We hypothesize that after duplication, *SoxE* genes underwent independent adaptation and specialization in the agnathan and gnathostome lineages, with divergent lamprey *SoxE* genes unable to compensate for loss of function among SoxE paralogs, whereas in vertebrates such as the frog, functional redundancy between *Sox9* and *Sox10* remains (Taylor and Labonne, 2005). Studies in other gnathostome vertebrates are also instructive. For example, *Sox10* mutants in both mouse and zebrafish have melanocyte, but not cartilage, defects, while Sox9 has now also been shown to possess a melanogenic role in zebrafish and humans (Greenhill et al., 2011; Passeron et al., 2007). These results suggest that gnathostome *Sox9* and *Sox10* may have overlapping functions in some

contexts but not others. However, the degree to which overlap in expression affects conserved protein function remains unclear. Here, we explore how functional specialization of SoxE proteins in lamprey, and functional redundancy among gnathostomes, may be related to early non-overlapping and overlapping expression domains of the *SoxE* genes in lampreys and gnathostomes respectively. Our results highlight the role of subfunctionalization in the diversification of neural crest derivatives, and reinforce the idea that the co-option and subsequent subfunctionalization of neural crest regulators, rather than the evolution of new protein functions, was key to diversification of the neural crest (Meulemans and Bronner-Fraser, 2004).

Materials and methods

Phylogenetic analysis of chordate SoxE proteins

ClustalX2.1 (Larkin et al., 2007) was used to align the following sequences:AAH85619 (MmSox8), AAH23808.1 (MmSox9), AAH23356.1 (MmSox10), AAX73357.1 (Drsox8), AAG09814.1 (Drsox9a), AAH67133.1 (Drsox9b), AAK84872.1 (Drsox10), AAQ67212.1 (XlSox8), AAI70060.1 (XlSox9), AAO13216.1 (XlSox10), AAF73917.1 (GgSox8), BAA25296.1 (GgSox9), AAD38050.2 (GgSox10), AAW34332.1 (PmSoxE1), ABC58684.1 (PmSoxE2), ABC58685.1 (PmSoxE3). *Branchiostoma floridae* SoxE sequence was obtained from cDNA clone bfne111n21 in Yu et al. (Yu et al., 2008).

Bioedit 7.0.9.0 (Hall, 1999b) was used to manually complete sequence alignments following ClustalX2.1 analysis. Maximum-Likelihood (ML), Minimum-Evolution (ME), and Neighbor-Joining (NJ) (Saitou and Nei, 1987) phylogenetic trees were constructed from the

aligned sequences with MEGA5 (Tamura et al., 2011), using the James-Taylor-Thornton (JTT) model with 1000 bootstrap replications and partial deletion with 50% site coverage cutoff.

Construction of amphioxus, zebrafish, and lamprey SoxE expression vectors

Sea lamprey and zebrafish (*Danio rerio*) SoxE full length coding sequences (AY830453, DQ328983, DQ328984, AF402677.1, and AF277096.1) were inserted into pCS2+ CMV promoter driven expression vectors, between EcoRI and XhoI (*PmSoxE1/PmSoxE2/ Drsox9a*), XhoI and XbaI (*PmSoxE3*), and ClaI and XhoI (*Drsox10*) sites of the multiple cloning site. SoxE-containing clones were diluted (35 ~ 55 ng/µl) in 0.1M KCl. Experiments using the Tol2 vector were conducted in the laboratories of DMM and RK. Following the excision of eGFP from the original Tol2 vector (T2KXIG) (Kawakami, 2004), amphioxus (*Branchiostoma floridae*) *SoxE,* and lamprey *PmSoxE1* and *PmSoxE2* coding sequences were subcloned into T2KXIG (Kawakami, 2004).

In situ hybridization

Lamprey *SoxE1*, *SoxE2,* and *SoxE3* gene fragments were cloned into the pGEM-Teasy vector using the following primer sets (*PmSoxE1* F: 5'-ACCTGCACAACGCCGAGCTG-3'; *PmSoxE1* R: 5'-CATGTCCACGTTGCTGAAGT-3'; *PmSoxE2* F: 5'-CGAGTTCGACCAGTACCTGCCC-3'; *PmSoxE2* R: 5'-ATGGTGGTGATGGTGGTGCTC-3'; *PmSoxE3* F: 5'-TGCTGGACGGCGGGGTGGTATTC-3'; *PmSoxE3* R: 5'-ACGTCCGCGCTGGGTGAGTCC. Sense and anti-sense digoxigenin-labeled (Roche) riboprobes were synthesized using T7 or Sp6 RNA polymerase. Zebrafish embryos were fixed in 4% paraformaldehyde (PFA) overnight at 4°C, washed twice in phosphate buffered saline (PBS), bleached in 3% hydrogen peroxide and 1% KOH in dH_2O, then stored in 100% methanol

(-20°C) prior to use. Whole-mount *in situ* hybridization was performed as described (Westerfield, 2007) but with the following modifications: embryos were incubated at 65°C for hybridization, followed by 50% formamide/2XSSCT washes (2x 30 min. at 65°C), then a single 2XSSCT wash (15 min. at 65°C).

Microinjection and fixation of zebrafish embryos

Zebrafish *colourless* (*cls*^{m618}) mutants were provided by RNK (Kelsh et al., 1996). Zygotes were injected through the chorion at the boundary between the yolk and blastomere, or directly into the blastomere. Injection of SoxE sequences was titered to 105~350 pg per embryo (pCS2+) or in T2KXIG as described (Kawakami, 2004). *cls* embryos injected with foreign constructs were reared to 96hpf, anesthetized using MS-222, and mounted on microscope slides for live imaging of pigmentation, then subsequently fixed in 2% trichloroacetic acid (TCA) prior to immunostaining.

Immunostaining

Mouse monoclonal antibodies against anti-human neuronal protein HuC/HuD (anti-HuC/D) (Invitrogen) were reconstituted in 500µl of (PBS, pH 7.4) containing 1% bovine serum albumin (BSA). For whole mount immunostaining, embryos were fixed in 2% TCA for 3 hours at room temperature (RT), washed in PBS, and PBT (PBS with 1% Triton-x-100), blocked with 10% goat serum, 1% BSA (4 hours at RT), incubated with anti-HuC/D (1:100) overnight at 4° C, washed 10 x 30 minutes in PBT, then incubated in Alexa Fluor 488 (Invitrogen) goat anti-mouse antibody (1:750) overnight at 4° C. Antibodies were resuspended in 1% goat serum, 1% BSA, and 1x PBT). Following removal of the secondary antibody, embryos were washed 10 x 30 min with PBT, and cleared in 30% glycerol at 4° C. Z–series image stacks were photographed on a

Zeiss AxioimagerZ1 compound microscope equipped with the Apotome optical sectioning module. Maximum intensity projections (MIP) were created using the Inside 4D module of the Zeiss Axiovision software package (v4.8.1).

Methylene Blue Staining

To determine the presence of xanthophores, embryos were stained in a solution of methylene blue as described (Le Guyader and Jesuthasan, 2002). Briefly, a 10% stock solution was made by dissolving methylene blue powder in deionized (dd)H_2O. A 1:20 working solution was made immediately prior to use using zebrafish system water. Embryos were stained for 24 hrs, rinsed to remove excess methylene blue, and then imaged using a Zeiss v12 stereomicroscope or Axioimager Z1 equipped with a Zeiss Axiocam MRc5 camera to determine the presence and morphology of xanthophores.

Reverse transcription PCR

Zebrafish wildtype (control) and *cls* mutant embryos were processed under identical conditions. Since *cls* embryos exhibit pigmentation in the retinal pigment epithelium (RPE) (Dutton et al., 2001), cranial elements anterior to the hindbrain were removed prior to RNA extraction. Injected embryos were reared to 96hpf, anesthetized in MS-222, decapitated, preserved in RNA*later* (n=10 per 1.5 ml tube) and stored at 4° C overnight prior to being transferred to -80° C for storage. Tissues in 1.5 ml Eppendorf tubes were snap frozen in liquid nitrogen and homogenized mechanically using a hand operated pellet pestle (Fisher Scientific) in conjunction with QIAshredders (Qiagen). RNA was extracted using the RNeasy kit (Qiagen) with on-column DNaseI treatment. Total RNA was diluted (1:10 ~ 1:30), treated with RNase-free DNaseI (Qiagen) and RNase inhibitor (Roche), then stored at -80° C. Total RNA was

reverse transcribed into cDNA using random hexamers or oligo-dT primers supplied by the Applied Biosystems Retroscript kit. Reactions were normalized across samples prior to PCR amplification using *β-actin* and *mitfa* specific primers. Primer designs with minimal conflicting targets were conducted using NCBI Primer-BLAST (Altschul et al., 1990): *mitfa* F: 5'-CGAGCCGGGGGTCTACGACA-3', *mitfa* R: 5'-GGAGGACAACAGCGGGTCGC-3', *β-actin* F:5'-GGTATGGGACAGAAAGACAG-3', *β-actin* R: 5'-AGAGTCCATCACGATACCAG-3'.

Statistical Analysis

Melanocyte and enteric neuron cell numbers were analyzed using one-way ANOVA with a significance threshold of $P<0.05$, followed by Games-Howell or Tukey-Kramer minimum significant difference (MSD) multiple-comparison *post-hoc* tests. Levene's test was used to evaluate homogeneity of variances.

Experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee at the University of Oklahoma.

Results

Comparative phylogenetic analysis of chordate SoxE protein sequences

A previous phylogeny based on a block alignment of conserved protein domains identified *PmSoxE3* as the lamprey ortholog to *Sox9* but identities of *PmSoxE1* and *PmSoxE2* remained uncertain (McCauley and Bronner-Fraser, 2006; Zhang et al., 2006). To clarify the identities of lamprey SoxE genes, we conducted a full length sequence alignment that included non-conserved regions outside the HMG, K2, and transactivation domains. Our results suggest

orthology of lamprey *PmSoxE2* to gnathostome *Sox10*, and reconfirm the orthology of *PmSoxE3* to gnathostome *Sox9* (Fig. 1; Supplementary Figs. S1, S2). Orthology of *PmSoxE1* remains unclear, still positioned basal to gnathostome *Sox8, Sox9,* or *Sox10* genes (Fig. 1), as previously described (McCauley and Bronner-Fraser, 2006).

Lamprey SoxE genes can induce differentiation of pigment cells in *cls* mutant zebrafish

To examine functional roles of SoxE proteins across the agnathan-gnathostome boundary, we expressed lamprey SoxE sequences in developing zebrafish embryos. In zebrafish, *sox10* is important for the specification of individual cell fates such as melanocytes. Moreover, neural crest cells that fail to undergo fate specification eventually die (Dutton et al., 2001). In lampreys, the knockdown of *PmSoxE2* led to abolished pigmentation, while *PmSoxE1* knockdown led to reduction in the size of pigment cells and loss of stellate morphology (Lakiza et al., 2011). In addition, *PmSoxE2* is expressed in lamprey dorsal root ganglia (Lakiza et al., 2011).

All three lamprey *SoxE* genes induced differentiation of melanophores when misexpressed in *cls* mutants, but with varying degrees of efficacy (Fig. 2, Table. 1). *PmSoxE2* had the greatest effect on melanophore differentiation with an average of 55 and a maximum of 171 pigment cells (n=114), compared to *PmSoxE1* (mean = 2.22, max = 8; n=68) and *PmSoxE3* (mean=4.79, max = 27; n=97) (Fig. 2G, comparison of *PmSoxE2* to *PmSoxE1* or *PmSoxE3* resulted in p-values < 0.05). Interestingly, the ability of *PmSoxE2* to promote melanogenesis exceeded that of the positive control (Fig. 2G, p-value < 0.05); injection of *Drsox10* into the *cls* mutant was able to promote differentiation of an average of 26.5 melanocytes, and a maximum number of 111 melanocytes (Fig. 2G). These differential effects on melanogenic rescue

following injection of each lamprey SoxE sequence are not likely to result from differences in levels of ectopic expression: *in situ* hybridization of lamprey constructs in zebrafish embryos showed comparable expression levels (Fig. 2H-J). The efficacy of plasmid injections into mutant embryos is summarized in Table 1. Both *Drsox10* and *PmSoxE2* were able to rescue melanogenesis in more than 93% of embryos, compared to *PmSoxE1* (25%) and *PmSoxE3* (30%). Interestingly, misexpression of *PmSoxE2* in wildtype zebrafish embryos resulted in excess pigmentation, especially in the trunk region (Supplementary Information Fig. S3). These results highlight the differential ability of lamprey protein sequences to regulate melanogenesis.

We also tested gnathostome SoxE sequences for differential melanogenic activity (Fig. 3). Neither zebrafish *sox9* ortholog (*sox9a* and *sox9b*) is expressed in differentiated melanocytes (Greenhill et al., 2011), while mutants in either gene do not show melanogenic defects (Yan et al., 2005). This suggests *Sox9* is unlikely to play a major melanogenic role, although Greenhill and colleagues noted melanogenic defects in *Sox9b* morphants on a sensitized background lacking *Sox10* activity (Greenhill et al., 2011). As expected, Sox10 (*XlSox10*) was able to rescue melanogenesis, but unexpectedly, *Drsox9a* promoted both a higher average and maximum number of differentiated melanocytes than *Xlsox10* (Fig. 3F), whereas, *Drsox8*, *XlSox9*, and *Drsox9b* were less efficient in their abilities to rescue melanogenesis (Tukey-Kramer minimum significant differences (MSD), observed > MSD). Induction of melanogenesis in *cls* mutants was comparable among embryos injected with zebrafish *DrSox10* inserted into CS2+ (Fig. 2) or other SoxE constructs cloned into Tol2 (Fig. 3).

Since vector-specific activity may account for observed functional differences in agnathan SoxE sequences cloned into pCS2 (Fig. 2C-E), versus robust recovery of melanogenesis by all gnathostome SoxE gene sequences cloned into Tol2 (Fig. 3A-E), we

inserted *PmSoxE1* and *PmSoxE2* coding sequences into both vectors to test for vector-specific melanogenic activity. As expected, *PmSoxE1* and *PmSoxE2* insertion into Tol2 promoted more robust melanogenic activity than insertion into pCS2 (compare Fig. 2C, D with Fig. 4D, G respectively). However, differences in SoxE activities that we observed using *pCS2-PmSoxE* constructs were validated by expression of *Tol2-PmSoxE1* and *Tol2-PmSoxE2* (Fig. 4, Table 2). *cls* $^{m618-/-}$ mutants were confirmed by examination of eyes for presence of iridophores (arrows in Fig. 4 G, H). Because of near total recovery of melanogenic activity in many mutants injected with *Tol2-PmSoxE2*, we were unable to determine the number of melanoblasts present in all rescued *cls* mutants. Thus, pigmentation recovery phenotypes were binned into 5 categories (I-V) (Fig. 4 and Table 2): I, few melanoblasts located along the dorsal head; II, melanoblasts along the dorsal head and ventrally just above the yolk; III, same as II and a small number of melanoblasts along the trunk; IV, moderate pigmentation along head and trunk; V, pigmentation indistinguishable from wildtype or heterozygous *m618* mutants. For *Tol2-PmSoxE1*-injected embryos, 69/113 *cls*$^{-/-}$ embryos were observed with melanoblasts (61% rescue) while 75/78 identified *cls*$^{-/-}$ embryos injected with *Tol2-PmSoxE2* contained differentiated melanoblasts (96% rescue). We were able to confirm only 78 *cls*$^{-/-}$ embryos by iridophore examination despite the expected ratio of 99/394 predicted by single allele Mendelian genetics (Table 2). This suggests that ~21/99 (21%) *cls* mutants could not be distinguished from the wildtype *Sox10* phenotype as a result of *PmSoxE2* integration and further highlights functional differences in melanogenic activities of *PmSoxE1* and *PmSoxE2* (Table 2). Taken together, data presented in Figs. 2-4 indicate a similar pattern in which Sox10/SoxE2 promote greater melanogenic activity than Sox8/Sox9/SoxE1/SoxE3, with the notable exception that *DrSox9a* was able to promote melanogenic activity equal to or exceeding *XlSox10*.

Lamprey *PmSoxE2* can regulate expression of *mitfa* in *cls* mutant zebrafish

PmSoxE2-induced melanogenesis suggests lamprey SoxE genes can regulate the melanogenic pathway in zebrafish. Thus, we examined expression of microphthalmia associated transcription factor a (*mitfa*), shown to be both necessary for melanocyte specification and differentiation and a direct target of Sox10 (Elworthy et al., 2003; Lister et al., 2001). *mitfa* is expressed in melanophores derived from neural crest cells, whereas *mitfb* is expressed in the RPE (Lister et al., 2001). Reverse transcription (rt)PCR, using RNA isolated from *PmSoxE2*-injected *cls* homozygous mutants, indicate *PmSoxE2* is able to drive *mitfa* expression in *cls* mutants (Fig. 2F).

Lamprey *PmSoxE2* can induce differentiation of xanthophores in *cls* mutant zebrafish

In addition to missing neural crest-derived melanophores, zebrafish *cls* mutants lack xanthophores, the cells responsible for the yellow pigmentation of fishes (Fig. 5A). *PmSoxE2* rescued differentiation of methylene blue-positive xanthophores (Fig. 5B – C). The morphology of these cells was similar to the stellate morphology of xanthophores in wildtype animals (Fig. 5D) (Le Guyader and Jesuthasan, 2002). *PmSoxE1* and *PmSoxE3* expression in *cls* mutants induced differentiation of only one to two xanthophores that were also of similar morphology to xanthophores in wildtype zebrafish (data not shown). Overall, these results indicate differential abilities of lamprey SoxE proteins to control chromogenic differentiation; *PmSoxE2* is able to regulate robust melanogenic and xanthogenic activity, whereas *PmSoxE1* and *PmSoxE3* are both limited in these activities.

Lamprey SoxE genes can induce differentiation of enteric neurons in *cls* mutants

Zebrafish *cls* mutants lack neurons of the enteric nervous system (ENS) (compare Fig. 6A and 6C) and have reduced numbers of sensory dorsal root ganglia (DRG) in the PNS (Carney et al., 2006; Elworthy et al., 2005). The variable baseline number of DRGs in *cls* mutant embryos made it difficult to assess rescue effects; therefore we focused on the rescue of enteric ganglia. ENS neurons were counted along the length of the gut immediately rostral to the anus. Uninjected (Fig. 6A) and mock injected wildtype embryos (Fig. 6B) both contain Hu-positive enteric neurons along and around the periphery of the gut in excess of 100 cells (Fig. 6A, B). Mock injected *cls* mutants still lacked enteric ganglia (Fig. 6C, G) (Carney et al., 2006; Elworthy et al., 2005). However, *PmSoxE2* was able to induce the formation of enteric ganglia with an average of 5.6 Hu-positive neurons per embryo, and a maximum of 36 (n=32; Fig. 6E, G), compared to *PmSoxE1* (mean = 1.3, max = 3, n=18; Fig. 6D, G). Interestingly, the maximum number of ENS neurons present following *PmSoxE2* injection (max = 36) exceeded that of the *Drsox10* control (mean = 4.1, max = 14, n=52), but a comparison of means revealed no statistical significance between the two groups (p-value = 0.357). Moreover, the ability of *PmSoxE3* (mean = 3.9, max = 24, n=44; Fig. 6F, G) to promote ENS rescue equaled that of the *Drsox10* positive control with no statistical difference seen between their means (p-value = 0.822). The following comparisons yielded p-values < 0.05: *Drsox10* / *PmSoxE1*, *PmSoxE1* / *PmSoxE2*, and *PmSoxE1* / *PmSoxE3*. The efficacy of plasmid injections into mutant embryos is summarized in Table 3.

Amphioxus *SoxE* can induce differentiation of melanophores and enteric neurons

Finally, we examined the ability of amphioxus (*Branchiostoma floridae*) SoxE to promote development of *sox10*-dependent neural crest derivatives. Misexpression of amphioxus *SoxE* in zebrafish *cls* mutants induced differentiation of both melanophores and enteric neurons in *cls* mutants (Fig. 7). Our results suggest that the ability of SoxE genes to regulate the

transcription of genes essential for melanogenic (e.g., *mitfa*) and enteric neurogenic (e.g., *phox2b*) activity likely predated the duplication and divergence of SoxE genes as well as appearance of the neural crest (Cossais et al., 2010a).

Discussion

Differential adaptation of SoxE proteins across the agnathan-gnathostome boundary

Lampreys diverged from gnathostome vertebrates over 450 million years ago (Janvier, 1996), and due to the independent evolution of these groups, much of the phylogenetic signal among SoxE genes has been lost in lampreys (McCauley and Bronner-Fraser, 2006). Previous results suggested both conservation and agnathan-specific specialization of SoxE genes (Lakiza et al., 2011), prompting further examination of lamprey SoxE gene evolution in the context of neural crest differentiation.

Our full-length phylogenetic analysis now places *PmSoxE2* in the vertebrate *Sox10* clade (Fig. 1, Supplementary information Fig. S1A). Heterospecific SoxE expression also suggests *PmSoxE2* is functionally similar to zebrafish *sox10*. In particular, our results show that chromogenic and enteric neurogenic neural crest roles of *PmSoxE2* and zebrafish *sox10* are conserved (Figs. 2, 6).

It is important to acknowledge that the current analysis is limited to the heterospecific expression of lamprey SoxE constructs in *sox10* (*cls*) mutants, and excludes analyses in *sox9* (*jellyfish, jef*) mutants lacking cartilage. While positive control injection of *Drsox10* into *cls* mutants is able to rescue melanogenic and enteric neurogenic activities, in our hands *DrSox9a* injection was unable to rescue the chondrogenic loss of function phenotype in zebrafish *Sox9a*

(*jef*) mutants, with the exception of a few small cartilage nodules (data not shown). Thus, we restricted this functional analysis to activity of lamprey SoxE genes in *Sox10* mutant embryos. Premigratory NCCs form in *cls* mutants but those fated to differentiate as pigment cells, enteric neurons, and dorsal root ganglia subsequently undergo apoptosis due to the failure of specification for these derivatives (Carney et al., 2006; Dutton et al., 2001; Kelsh and Raible, 2002). These observations suggest that early in vertebrate evolution, *Sox10* may have acquired roles specific to regulation of pigment cell and neuronal specification since *PmSoxE2* is able to induce differentiation of melanophores, xanthophores, and enteric neurons in *cls* mutants (Figs. 2, 4-6), but may be unable to induce cartilage formation in *jef* mutants (data not shown). Furthermore, heterospecific expression of *PmSoxE2* in wildtype zebrafish causes formation of excess melanophores (Supplementary Information, Fig. S3). These results highlight both sequence and functional similarities between lamprey *PmSoxE2* and zebrafish *sox10*.

Functional redundancy among gnathostome SoxE genes has been reported (Taylor and Labonne, 2005). To determine if lamprey SoxE proteins are functionally equivalent, we examined melanogenic and enteric neurogenic activity among the lamprey SoxE genes in a zebrafish host lacking Sox10. *PmSoxE2* was able to promote melanogenesis to a greater extent than either *PmSoxE1* or *PmSoxE3* (Fig. 2), whereas frog and zebrafish SoxE constructs were all able to promote robust melanogenic activity (Fig. 3). For neurogenic functions, the ability of both *PmSoxE2* and *PmSoxE3* to promote the differentiation of enteric neurons was comparable to *Drsox10* (Fig. 6). In comparison, *PmSoxE1*does not show any significant enteric neurogenic activity (Fig. 6), instead being required for chondrogenesis and suggesting functional specialization of lamprey SoxE genes following their duplication (Lakiza et al., 2011; McCauley and Bronner-Fraser, 2006). Furthermore, in contrast to the ability of *PmSoxE2* to promote a

greater number of melanophores than *Drsox10* (Fig. 2), its enteric neurogenic activity did not exceed that of *Drsox10* (Fig. 6).

Differential abilities among the three lamprey SoxE genes to direct pigment cell differentiation in the zebrafish *cls* background may represent lamprey-specific functional adaptations among these duplicated genes. Alternatively, it is possible that differential translational efficiencies may account for these observed differences. However, the ability of *PmSoxE3* to induce enteric neuron differentiation at levels comparable to *Drsox10* and *PmSoxE2* suggests that it is functional in *cls* mutants (Fig. 6G). *Sox10* function also requires protein domains in addition to the highly conserved HMG box, including the DNA-dependent dimerization domain (Peirano et al., 2000; Peirano and Wegner, 2000), the K2 domain (Schepers et al., 2000; Wegner, 1999), and the C-terminal transactivation domain (Cossais et al., 2010b). Earlier studies of *in vitro* transfection assays showed that human SOX10 and PAX3 synergistically activate *MITF* (Bondurand et al., 2000; Potterf et al., 2000). Differentiation of melanocytes and enteric neurons require functional Sox10 dimerization and K2 domains in mice, suggesting that Sox10 may act in conjunction with other protein partners to activate distal enhancers of target genes (Schreiner et al., 2007), as shown for activation of *Mitf* (Watanabe et al., 2002) (e.g., *phox2b* and others). Future studies to characterize lamprey SoxE domains will be useful for understanding how SoxE proteins have diverged in function throughout evolution.

Individuation of gene expression is permissive for protein subfunctionalization

Mutations in *cis*-regulatory sequences are used to explain how genes can undergo subfunctionalization following duplication events. However, there have been few studies that

provide a clear link between the subfunctionalization of protein activity and the expression differences resulting from *cis*-regulatory changes. Indeed, the duplication, degeneration, complementation model for gene evolution focuses on the role of degenerative regulatory elements in duplicated genes to preserve these duplicates (Force et al., 1999), but the mechanisms that regulate the evolution of protein activity among duplicates are less well understood.

The SoxE paralogs, *Sox8*, *9*, and *10* are considered to be functionally redundant in some gnathostomes (Taylor and Labonne, 2005). Here, we have shown that duplicated SoxE proteins in the agnathan sea lamprey are not equally able to rescue melanogenesis or differentiation of enteric neurons when expressed in a heterospecific *cls* mutant zebrafish background. This work suggests functional differences in SoxE genes have arisen along the lamprey lineage, but such differences may be less obvious among gnathostome SoxE genes. An analysis of SoxE gene expression profiles is instructive and suggests that overlapping expression may generate constraints on the evolution of gnathostome SoxE proteins to prevent their functional divergence while such constraints may not be present in the lamprey where early expression domains overlap to a lesser extent than in gnathostomes (Fig. 8; adapted from Aoki et al., 2003; Lakiza et al., 2011; O'Donnell et al., 2006; Sauka-Spengler et al., 2007; Spokony et al., 2002). For example, all three SoxE paralogs are expressed in an overlapping pattern during neural crest specification (stage 16) and cranial NCC migration in the *Xenopus* embryo (stage 25) (Fig. 8A, B). A similar comparison of expression among lamprey SoxE paralogs reveals that even at the earliest stage of SoxE expression in premigratory neural crest (Tahara Stage 21) there are already non-overlapping expression domains among the SoxE paralogs (Fig. 8C); at Tahara stage 22 (Fig. 8D), lamprey SoxE genes are expressed throughout premigratory cranial neural crest, but

only *PmSoxE2* is expressed in migratory NCCs invading the pharyngeal region, or in trunk NCCs along the anteroposterior axis.

These comparisons support a scenario in which constraints on evolution of SoxE function in jawed vertebrates may be related to overlapping expression, where loss of expression by any SoxE gene could be compensated for by either of the remaining cognates. This would require that all members of the SoxE subfamily retain overlapping functional activity. Subsequent non-overlapping activity resulting from subfunctionalization and/or neofunctionalization would be dependent on changes in *cis*-regulatory sequences to drive the functional specialization of each gnathostome SoxE protein.

In contrast, lampreys, having evolved independently from gnathostomes for half a billion years, appear to have used a separate strategy to drive the functional specialization of SoxE genes. Non-overlapping expression during early stages of embryogenesis might have led to a release from constraints that prevented diversification of function still seen in jawed vertebrates. Relaxation from these constraints would have enabled SoxE paralogs in lampreys to acquire functional specialization based on accumulation of mutations in SoxE amino acid sequences, in addition to the *cis*-regulatory changes that drove expression differences. Thus, differential gene expression led to evolution of protein function among gene duplicates which, over time, have lost the ability to compensate for the loss of redundant functions by duplicated paralogs, whereas the early overlapping expression of gnathostome SoxE cognates during NCC specification and migration correlates with redundant functional activity.

A simple test of this proposal would be to replace each SoxE coding sequence (e.g., *Sox10*) with one of its paralogs (e.g., *Sox9*) to determine the ability of each protein to

compensate for loss of the replaced gene sequence. Since the inserted gene would be under control of the original (i.e., *Sox10*) regulatory sequence, one might expect to uncover functional differences that could not be determined without the precise spatiotemporal regulation of expression. Indeed, Wegner's group found that the *Drosophila* SoxE ortholog *Sox100b,* inserted in place of mouse *Sox10*, was able to direct aspects of neural crest and oligodendrocyte development, suggesting co-option of SoxE proteins in vertebrate gene regulatory networks. Interestingly, defects in late-developing neural crest lineages suggested SoxE proteins have also undergone functional specializations (Cossais et al., 2010a).

The idea that developmental expression differences may drive functional changes among gene duplicates, while overlapping expression may constrain it, is supported by studies examining the developmental roles of duplicated *tfap2* genes in vertebrates. Five *tfap2* paralogs exist in vertebrates, having duplicated from a single ancestral *tfap2* gene (Hoffman et al., 2007; Meulemans and Bronner-Fraser, 2002). Four *tfap2* paralogs (*tfap2a,b,c,d*) likely arose from a single vertebrate *tfap2* copy and are all functional in early ectodermal and neural crest development. *tfap2a* and *tfap2c* are functionally redundant in ectodermal and neural crest development (Hoffman et al., 2007; Li and Cornell, 2007) and *tfap2a, tfap2b, tfap2c* and *tfap2e* are all expressed in partially overlapping patterns in brain (Van Otterloo et al., 2012). Moreover, injection of mRNAs encoding *tfap2a, tfap2b, tfap2c*, or *tfap2e* were each able to rescue expression of neural crest markers in a *tfap2a/c* deficient zebrafish background. On the other hand, the remaining paralog, *tfap2d*, diverged early in vertebrate evolution and is least similar at the amino acid level among all *tfap2* paralogs (Hoffman et al., 2007). *tfap2d* also appears to play no role in early development. Additionally, injection of mRNA encoding *tfap2d* was unable to rescue neural crest marker expression in *tfap2a/c* deficient zebrafish (Van Otterloo et al., 2012).

These observations support the hypothesis that the accumulation of developmental expression differences may be pivotal for driving the evolution of protein function.

Melanogenic, neurogenic, and chondrogenic functions of *Sox8*, *Sox9*, and *Sox10* predate duplication

Two rounds of gene duplication events occurred early in vertebrate evolution (Ohno, 1970; Wada and Makabe, 2006). It has been proposed that these gene duplications drove the evolution of vertebrate specific features by serving as gene repertoires (Force et al., 1999; Kasahara, 2007; Kuraku et al., 2009; Lynch and Force, 2000; Lynch et al., 2006; Meyer and Van de Peer, 2005; Ohno, 1970; Wagner et al., 2003; Zhang, 2003). Consistent with this notion, transcription factors important for neural crest specification often exist as duplicate copies in vertebrates (Guth and Wegner, 2008). Furthermore, comparison of duplicated genes originating from ancient duplication events among organisms may be used to distinguish vertebrates from invertebrates at the genomic level (Kuraku et al., 2009). The presence of migratory neural crest-like cells in several invertebrate groups suggests the neural crest likely arose from a pre-existing migratory population of cells (Abitua et al., 2012; Jeffery, 2007; Jeffery et al., 2008; Jeffery et al., 2004; Medeiros, 2013). A comparison of the neural crest gene regulatory networks (GRNs) among basal chordates and vertebrates also reveals the pre-existence of fundamental components of the GRN in early chordates, which may have been co-opted by the vertebrate common ancestor to give rise to the neural crest (Green et al., 2015; Medeiros and Crump, 2012; Meulemans and Bronner-Fraser, 2004; Meulemans et al., 2003; Sauka-Spengler et al., 2007; Stone and Hall, 2004). This raises the question of whether the NCC-specific functions of

vertebrate *Sox8*, *Sox9*, and *Sox10* arose through neofunctionalization following gene duplication, or whether they collectively possess the functionality of the ancestral vertebrate *SoxE* (Van Otterloo et al., 2012). Our data, taken together with recent findings in amphioxus (Jandzik et al., 2015), suggest the latter scenario, i.e., that the proto-vertebrate *SoxE* already possessed melanogenic, neurogenic, and chondrogenic regulatory capabilities prior to gene duplication. We show that amphioxus *SoxE* induces differentiation of melanophores and enteric neurons (Fig. 7). Additionally, amphioxus *SoxE* causes an expansion in the expression of the NC marker *Slug* when injected into *Xenopus* embryos suggesting regulatory activity of SoxE that might have been co-opted by the neural crest in the vertebrate ancestor (Supplementary Fig. 5).

Conclusions

In sum, our results reveal a complex history of SoxE duplication and subfunctionalization in vertebrates. Based on sequence analysis and our functional investigation of SoxE genes near the base of vertebrates, we suggest that the melanogenic, neurogenic, and chondrogenic (Jandzik et al., 2015) capabilities of the ancestral SoxE gene were co-opted by the neural crest in the ancestral vertebrate. We suggest that early duplication of the single *SoxE* gene into *Sox9* and *Sox10* clades occurred in the common ancestor of jawed and jawless vertebrates. Following this event, NCC-specific neurogenic and melanogenic functions were partitioned to *Sox10,* while *Sox9* acquired roles in chondrogenic specification and morphogenesis. In gnathostomes the chondrogenic specification and morphogenetic roles were maintained by *Sox9*. However, in lampreys, *PmSoxE3* lost the early cartilage specification function, instead maintaining only its later role in cartilage morphogenesis (Lakiza et al., 2011). This was made possible by the appearance of the agnathan-specific SoxE paralog, *PmSoxE1*, which maintains the cartilage specification role in lampreys (Lakiza et al., 2011; McCauley and Bronner-Fraser, 2006).

Our data establish the early appearance of the *Sox9* and *Sox10* clades, but the timing and nature of the duplications that generated lamprey *PmSoxE1* and gnathostome *Sox8* are less clear. In gnathostomes, *Sox8* groups with *Sox9*, albeit at low confidence values (Fig. 1) (Meulemans and Bronner-Fraser, 2007). This could suggest that *Sox8* was generated from an ancestral *Sox8*/*9* by the second whole genome duplication in the gnathostome lineage. Presumably, a second *Sox10* paralog would have also been formed at this time, to be later lost in both gnathostomes and lampreys. If lampreys diverged from gnathostomes after this second duplication event, as has been proposed (Kuraku et al., 2009), it is conceivable that lamprey *PmSoxE1* could represent a highly divergent *Sox8* ortholog. Alternately, *PmSoxE1* may be a divergent lamprey-specific duplicate of *Sox9* or *Sox10*. This scenario would not require lampreys to have undergone the same whole-genome duplications as gnathostomes, or to have lost *Sox8*. With availability of the sea lamprey genome (Smith et al., 2013), it may now be possible to distinguish between these alternatives as improvements are made to the genome assembly. Such analyses will help to paint a more accurate picture of *SoxE* evolution as well as the evolution and diversification of neural crest cells.

Acknowledgments

We thank Dr. John Masly for assistance with statistical and phylogenetic analyses. This project was supported by NIH 1R03DE018159-01 and NSF IOS-0842763 grants to DWM, NSF IOS-0920751 to DMM, and funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 661616 (RDB). The study sponsors had no role in design, collection, analysis or interpretation of data. Experiments were conceived by EML, DWM, and DMM and conducted by EML, DWM, DMM, RDB, and RNK. Manuscript preparation was by EML, with assistance by DWM and DMM. TY, KN, and RDB conducted microinjection and microscopy analysis of results.

Figure Legends

Figure 1. Phylogenetic analysis of chordate SoxE genes, including those used in this study. A Neighbor-Joining (NJ) tree was constructed in MEGA5 using a Clustal alignment of full length amino acid sequences from chordate SoxE genes; James-Taylor-Thornton (JTT) model with 1000 bootstrap replications and partial deletion with 30 % site coverage cutoff. Phylogenies were calculated using NJ, Minimum-Evolution (ME), and Maximum-Likelihood (ML) methods. Numbers at each node represent bootstrap values for NJ, ME, and ML from top to bottom respectively.

Figure 2. Melanogenesis in *cls* mutants resulting from expression of lamprey SoxE sequences inserted into pCS2+ vectors. (A – E) 96 hpf zebrafish larvae. (A) wildtype. (B) *cls*; RPE, retinal pigment epithelium. Expression of *PmSoxE1* (C), *PmSoxE2* (D), and *PmSoxE3* (E) in *cls* mutants. (F) RT-PCR of *PmSoxE2*-injected *cls* mutants. *mitfa* expression in a *cls* embryo resulting from *PmSoxE2* injection (lane 1) is absent in the control *cls* embryo (lane 2). *β-actin* internal control (lanes 3 – 4). (G) box and whisker plot indicating the number of melanophores in zebrafish injected with *Drsox10* (positive control), lamprey *PmSoxE1*, *PmSoxE2*, and *PmSoxE3*, and *cls* embryos mock-injected with empty pCS2 vector as a negative control. Whiskers represent standard error. Each dot above the positive whiskers represents individual data points of the first quartile. Sample sizes are indicated under each bar along the X-axis. Comparison of means among samples all resulted in statistical significance (ANOVA, $p\text{-}value<0.05$). (H-J) *in situ* hybridization using lamprey SoxE probes in zebrafish wildtype embryos injected with either lamprey *SoxE1* (H), *SoxE2* (I), or *SoxE3* (J) reveals similar mRNA expression, suggesting lamprey *SoxE* constructs injected into zebrafish embryos yields a comparable pattern and expression level irrespective of the lamprey SoxE gene sequence. Results

shown are representative of 19/58 (*PmSoxE1*), 45/229 (*PmSoxE2*, and 90/356 (*PmSoxE3*) injections of lamprey constructs into early cleavage stage wildtype zebrafish embryos. vh, ventral horn; ys, yolk sac; ye, yolk extension; ds, dorsal stripe; ls, lateral stripe; vs, ventral stripe. Arrows point to melanophores, arrowheads point to iridophores. Orientation: anterior facing left.

Figure 3. Gnathostome *Sox8*, *Sox9*, and *Sox10* sequences inserted into Tol2 vectors can all promote melanogenesis in *cls* mutants. (A – E) 96 hpf zebrafish larvae. Expression of *Drsox8* (A), *Drsox9a* (B), *Drsox9b* (C), *XlSox9* (D), and *XlSox10* (E) in *cls* mutants. (F) box and whisker plot indicating the number of melanophores rescued. Whiskers represent standard error. Each dot above the positive whiskers represents individual data points of the first quartile. Sample sizes are indicated under each bar on the X-axis.

Figure 4. Melanogenesis in *cls* mutants resulting from expression of *PmSoxE1* and *PmSoxE2* sequences inserted into Tol2 vectors. (A – H) 96 hpf zebrafish larvae. Expression of *PmSoxE1* (A – D) and *PmSoxE2* (E – G) and wildtype melanogenic phenotype (H). (I – V) pigmentation rescue phenotypes categorized as described in the text. Arrows indicate the absence of iridophores in *cls* mutant larvae (G) versus their presence in wildtype siblings (H).

Figure 5. *PmSoxE2* promotes differentiation of xanthophores in *cls* mutant zebrafish. (A – B) 76 hpf zebrafish embryos stained with methylene blue to detect presence of xanthophores. (A) *cls* mutant lacking differentiated xanthophores. (B) xanthophores in *cls* mutants injected with the pCS2-*PmSoxE2* construct. (C) inset in "B" ; higher magnification highlights the presence of xanthophores following *PmSoxE2* injection. (D) morphology of xanthophores in a wildtype 76 hpf embryo. Orientation: anterior facing left.

Figure 6. pCS2-PmSoxE heterospecific expression promotes differentiation of enteric neurons along the larval gut in *cls* mutants. (A – F) anti-HuC/D immunofluorescence in 96 hpf zebrafish. (A) wildtype. (B) control mock-injected wildtype. (C) *cls* embryo mock-injected with empty pCS2 vector. Note the absence of HuC/D-positive enteric neurons in *cls* embryos. Enteric neurons are present in *cls* mutants injected with lamprey *PmSoxE1* (D), *PmSoxE2* (E), and *PmSoxE3* (F) constructs (arrows in D-F). (G) box and whisker plots indicate the number of enteric neurons present in zebrafish injected with *Drsox10, PmSoxE1*, *PmSoxE2*, and *PmSoxE3* constructs. Whiskers represent standard error. Each dot above the positive whiskers represents individual data points of the first quartile. Sample sizes are indicated under each bar on the X-axis. Arrows indicate anti-HuC/D positive enteric neurons. Orientation: anterior facing left.

Figure 7. Amphioxus *SoxE* promotes differentiation of melanophores and enteric neurons. (A) differentiated melanophores in 96 hpf *cls* mutants injected with a Tol2:*BfSoxE* construct. (B – D) anti-HuC/D immunofluoresence in 96 hpf zebrafish indicates the presence of enteric neurons. (B) control wildtype. (C) *cls* mutants lack HuC/D-positive enteric neurons. (D) *BfSoxE* promotes differentiation of enteric neurons in *cls* mutants. Orientation: anterior facing left.

Figure 8. Cartoon representation of SoxE gene expression in *Xenopus*, a representative gnathostome (A, B), and in lamprey (C, D). (A). *XlSox8*, *XlSox9*, and *XlSox10* expression domains largely overlap along the neural plate border (NPB) at stage 16 as neural crest specification occurs, but with a non-overlapping *XlSox8* domain present in the cement gland (CG). At stage 25 (B), *XlSox8*, *XlSox9*, and *XlSox10* are all expressed in neural crest cells migrating into the pharyngeal pouches (red arrowheads) and *XlSox9* expression in the eye (arrow in B). (C) Early SoxE expression in the stage 21 lamprey. *PmSoxE1*, *PmSoxE2* and *PmSoxE3* are co-expressed in a small region of premigratory neural crest (arrow in C), but non-overlapping

PmSoxE2 (blue) and *PmSoxE3* (salmon) expression is already apparent along the midbrain and at the otic vesicle respectively. At stage 22 (D), *PmSoxE1* (green) and *PmSoxE3* (salmon) are expressed in premigratory neural crest, but only *PmSoxE2* (blue) is expressed in the neural crest cells migrating toward the presumptive pharyngeal region (arrowheads in D) and also along the trunk. Schematic cartoon expression patterns are adapted from (Aoki et al., 2003; Kelsh et al., 1996; Lakiza et al., 2011; Sauka-Spengler et al., 2007; Spokony et al., 2002). Anterior faces left for all illustrations.

References

Abitua, P.B., Wagner, E., Navarrete, I.A., Levine, M., 2012. Identification of a rudimentary neural crest in a non-vertebrate chordate. Nature 492, 104-107.

Akiyama, H., Chaboissier, M.C., Martin, J.F., Schedl, A., de Crombrugghe, B., 2002. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Genes Dev 16, 2813-2828.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.

Aoki, Y., Saint-Germain, N., Gyda, M., Magner-Fink, E., Lee, Y.H., Credidio, C., Saint-Jeannet, J.P., 2003. Sox10 regulates the development of neural crest-derived melanocytes in Xenopus. Developmental biology 259, 19-33.

Bell, D.M., Leung, K.K., Wheatley, S.C., Ng, L.J., Zhou, S., Ling, K.W., Sham, M.H., Koopman, P., Tam, P.P., Cheah, K.S., 1997. SOX9 directly regulates the type-II collagen gene. Nat Genet 16, 174-178.

Bondurand, N., Kobetz, A., Pingault, V., Lemort, N., Encha-Razavi, F., Couly, G., Goerich, D.E., Wegner, M., Abitbol, M., Goossens, M., 1998. Expression of the SOX10 gene during human development. FEBS letters 432, 168-172.

Bondurand, N., Pingault, V., Goerich, D.E., Lemort, N., Sock, E., Le Caignec, C., Wegner, M., Goossens, M., 2000. Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. Hum Mol Genet 9, 1907-1917.

Bowles, J., Schepers, G., Koopman, P., 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Developmental biology 227, 239-255.

Britsch, S., Goerich, D.E., Riethmacher, D., Peirano, R.I., Rossner, M., Nave, K.A., Birchmeier, C., Wegner, M., 2001. The transcription factor Sox10 is a key regulator of peripheral glial development. Genes Dev 15, 66-78.

Carney, T.J., Dutton, K.A., Greenhill, E., Delfino-Machin, M., Dufourcq, P., Blader, P., Kelsh, R.N., 2006. A direct role for Sox10 in specification of neural crest-derived sensory neurons. Development 133, 4619-4630.

Cheng, Y., Cheung, M., Abu-Elmagd, M.M., Orme, A., Scotting, P.J., 2000. Chick sox10, a transcription factor expressed in both early neural crest cells and central nervous system. Brain research. Developmental brain research 121, 233-241.

Cheung, M., Briscoe, J., 2003. Neural crest development is regulated by the transcription factor Sox9. Development 130, 5681-5693.

Chiang, E.F., Pai, C.I., Wyatt, M., Yan, Y.L., Postlethwait, J., Chung, B., 2001. Two sox9 genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. Developmental biology 231, 149-163.

Cossais, F., Sock, E., Hornig, J., Schreiner, S., Kellerer, S., Bosl, M.R., Russell, S., Wegner, M., 2010a. Replacement of mouse Sox10 by the Drosophila ortholog Sox100B provides evidence for co-option of SoxE proteins into vertebrate-specific gene-regulatory networks through altered expression. Developmental biology 341, 267-281.

Cossais, F., Wahlbuhl, M., Kriesch, J., Wegner, M., 2010b. SOX10 structure-function analysis in the chicken neural tube reveals important insights into its role in human neurocristopathies. Hum Mol Genet 19, 2409-2420.

Cresko, W.A., Yan, Y.L., Baltrus, D.A., Amores, A., Singer, A., Rodriguez-Mari, A., Postlethwait, J.H., 2003. Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. Dev Dyn 228, 480-489.

Dutton, K.A., Pauliny, A., Lopes, S.S., Elworthy, S., Carney, T.J., Rauch, J., Geisler, R., Haffter, P., Kelsh, R.N., 2001. Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. Development 128, 4113-4125.

Elworthy, S., Lister, J.A., Carney, T.J., Raible, D.W., Kelsh, R.N., 2003. Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. Development 130, 2809-2818.

Elworthy, S., Pinto, J.P., Pettifer, A., Cancela, M.L., Kelsh, R.N., 2005. Phox2b function in the enteric nervous system is conserved in zebrafish and is sox10-dependent. Mech Dev 122, 659-669.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531-1545.

Gans, C., Northcutt, R.G., 1983. Neural Crest and the Origin of Vertebrates: A New Head. Science 220, 268-273.

Green, S.A., Simoes-Costa, M., Bronner, M.E., 2015. Evolution of vertebrates as viewed from the crest. Nature 520, 474-482.

Greenhill, E.R., Rocco, A., Vibert, L., Nikaido, M., Kelsh, R.N., 2011. An iterative genetic and dynamical modelling approach identifies novel features of the gene regulatory network underlying melanocyte development. PLoS Genet 7, e1002265.

Guth, S.I., Wegner, M., 2008. Having it both ways: Sox protein function between conservation and innovation. Cell Mol Life Sci 65, 3000-3018.

Hall, B.K., 1999a. The Neural Crest in Development and Evolution. Springer-Verlag, New York.

Hall, T.A., 1999b. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95-98.

Hoffman, T.L., Javier, A.L., Campeau, S.A., Knight, R.D., Schilling, T.F., 2007. Tfap2 transcription factors in zebrafish neural crest development and ectodermal evolution. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 308B, 679-691.

Holland, L.Z., Short, S., 2008. Gene duplication, co-option and recruitment during the origin of the vertebrate brain from the invertebrate chordate brain. Brain, behavior and evolution 72, 91-105.

Hong, C.S., Saint-Jeannet, J.P., 2005. Sox proteins and neural crest development. Semin Cell Dev Biol 16, 694-703.

Honore, S.M., Aybar, M.J., Mayor, R., 2003. Sox10 is required for the early development of the prospective neural crest in Xenopus embryos. Developmental biology 260, 79-96.

Jandzik, D., Garnett, A.T., Square, T.A., Cattell, M.V., Yu, J.-K., Medeiros, D.M., 2015. Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. Nature 518, 534-537.

Janvier, P., 1996. Early Vertebrates. Oxford University Press, Oxford, England.

Jeffery, W.R., 2007. Chordate ancestry of the neural crest: new insights from ascidians. Semin Cell Dev Biol 18, 481-491.

Jeffery, W.R., Chiba, T., Krajka, F.R., Deyts, C., Satoh, N., Joly, J.S., 2008. Trunk lateral cells are neural crest-like cells in the ascidian Ciona intestinalis: insights into the ancestry and evolution of the neural crest. Developmental biology 324, 152-160.

Jeffery, W.R., Strickler, A.G., Yamamoto, Y., 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431, 696-699.

Kasahara, M., 2007. The 2R hypothesis: an update. Curr Opin Immunol 19, 547-552.

Kawakami, K., 2004. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. Methods Cell Biol 77, 201-222.

Kellerer, S., Schreiner, S., Stolt, C.C., Scholz, S., Bosl, M.R., Wegner, M., 2006. Replacement of the Sox10 transcription factor by Sox8 reveals incomplete functional equivalence. Development 133, 2875-2886.

Kelsh, R.N., Brand, M., Jiang, Y.J., Heisenberg, C.P., Lin, S., Haffter, P., Odenthal, J., Mullins, M.C., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Kane, D.A., Warga, R.M., Beuchle, D., Vogelsang, L., Nusslein-Volhard, C., 1996. Zebrafish pigmentation mutations and the processes of neural crest development. Development 123, 369-389.

Kelsh, R.N., Eisen, J.S., 2000. The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. Development 127, 515-525.

Kelsh, R.N., Raible, D.W., 2002. Specification of zebrafish neural crest. Results Probl Cell Differ 40, 216-236.

Kondoh, H., Kamachi, Y., 2010. SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. Int J Biochem Cell Biol 42, 391-399.

Kuraku, S., Meyer, A., Kuratani, S., 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol Biol Evol 26, 47-59.

Kuratani, S., Kuraku, S., Murakami, Y., 2002. Lamprey as an evo-devo model: lessons from comparative embryology and molecular phylogenetics. Genesis 34, 175-183.

Lakiza, O., Miller, S., Bunce, A., Lee, E.M., McCauley, D.W., 2011. SoxE gene duplication and development of the lamprey branchial skeleton: Insights into development and evolution of the neural crest. Developmental biology 359, 149-161.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Laudet, V., Stehelin, D., Clevers, H., 1993. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res 21, 2493-2501.

Le Guyader, S., Jesuthasan, S., 2002. Analysis of xanthophore and pterinosome biogenesis in zebrafish using methylene blue and pteridine autofluorescence. Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society 15, 27-31.

LeDouarin, N.M., Kalchiem, C., 1999. The Neural Crest. Cambridge University Press, Cambridge, UK.

Li, M., Zhao, C., Wang, Y., Zhao, Z., Meng, A., 2002. Zebrafish sox9b is an early neural crest marker. Dev Genes Evol 212, 203-206.

Li, W., Cornell, R.A., 2007. Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. Developmental biology 304, 338-354.

Lister, J.A., Close, J., Raible, D.W., 2001. Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. Developmental biology 237, 333-344.

Lynch, M., Force, A., 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154, 459-473.

Lynch, V.J., Roth, J.J., Wagner, G.P., 2006. Adaptive evolution of Hox-gene homeodomains after cluster duplications. BMC Evol Biol 6, 86.

McCauley, D.W., Bronner-Fraser, M., 2006. Importance of SoxE in neural crest development and the evolution of the pharynx. Nature 441, 750-752.

Medeiros, D.M., 2013. The evolution of the neural crest: new perspectives from lamprey and invertebrate neural crest-like cells. Wiley Interdisciplinary Reviews-Developmental Biology 2, 1-15.

Medeiros, D.M., Crump, J.G., 2012. New perspectives on pharyngeal dorsoventral patterning in development and evolution of the vertebrate jaw. Developmental biology 371, 121-135.

Mertin, S., McDowall, S.G., Harley, V.R., 1999. The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic Acids Res 27, 1359-1364.

Meulemans, D., Bronner-Fraser, M., 2002. Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. Development 129, 4953-4962.

Meulemans, D., Bronner-Fraser, M., 2004. Gene-regulatory interactions in neural crest evolution and development. Dev Cell 7, 291-299.

Meulemans, D., Bronner-Fraser, M., 2007. Insights from amphioxus into the evolution of vertebrate cartilage. PloS one 2, e787.

Meulemans, D., McCauley, D., Bronner-Fraser, M., 2003. Id expression in amphioxus and lamprey highlights the role of gene cooption during neural crest evolution. Developmental biology 264, 430-442.

Meyer, A., Van de Peer, Y., 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27, 937-945.

O'Donnell, M., Hong, C.S., Huang, X., Delnicki, R.J., Saint-Jeannet, J.P., 2006. Functional analysis of Sox8 during neural crest development in Xenopus. Development 133, 3817-3826.

Ohno, S., 1970. Evolution by Gene Duplication, 1st ed. Springer-Verlag.

Passeron, T., Valencia, J.C., Bertolotto, C., Hoashi, T., Le Pape, E., Takahashi, K., Ballotti, R., Hearing, V.J., 2007. SOX9 is a key plaver in ultraviolet B-induced melanocyte differentiation

and pigmentation. Proceedings of the National Academy of Sciences of the United States of America 104, 13984-13989.

Peirano, R.I., Goerich, D.E., Riethmacher, D., Wegner, M., 2000. Protein zero gene expression is regulated by the glial transcription factor Sox10. Molecular and cellular biology 20, 3198-3209.

Peirano, R.I., Wegner, M., 2000. The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. Nucleic Acids Res 28, 3047-3055.

Pingault, V., Bondurand, N., Kuhlbrodt, K., Goerich, D.E., Prehu, M.O., Puliti, A., Herbarth, B., Hermans-Borgmeyer, I., Legius, E., Matthijs, G., Amiel, J., Lyonnet, S., Ceccherini, I., Romeo, G., Smith, J.C., Read, A.P., Wegner, M., Goossens, M., 1998. SOX10 mutations in patients with Waardenburg-Hirschsprung disease. Nat Genet 18, 171-173.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., Yan, Y.L., 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet 20, 481-490.

Potterf, S.B., Furumura, M., Dunn, K.J., Arnheiter, H., Pavan, W.J., 2000. Transcription factor hierarchy in Waardenburg syndrome: regulation of MITF expression by SOX10 and PAX3. Hum Genet 107, 1-6.

Prior, H.M., Walter, M.A., 1996. SOX genes: architects of development. Mol Med 2, 405-412.

Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I.T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.

Rau, M.J., Fischer, S., Neumann, C.J., 2006. Zebrafish Trap230/Med12 is required as a coactivator for Sox9-dependent neural crest, cartilage and ear development. Developmental biology 296, 83-93.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4, 406-425.

Sauka-Spengler, T., Meulemans, D., Jones, M., Bronner-Fraser, M., 2007. Ancient evolutionary origin of the neural crest gene regulatory network. Dev Cell 13, 405-420.

Schepers, G.E., Bullejos, M., Hosking, B.M., Koopman, P., 2000. Cloning and characterisation of the Sry-related transcription factor gene Sox8. Nucleic Acids Res 28, 1473-1480.

Schreiner, S., Cossais, F., Fischer, K., Scholz, S., Bosl, M.R., Holtmann, B., Sendtner, M., Wegner, M., 2007. Hypomorphic Sox10 alleles reveal novel protein functions and unravel developmental differences in glial lineages. Development 134, 3271-3281.

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., Morgan, J.R., Buxbaum, J.D., Sachidanandam, R., Sims, C., Garruss, A.S., Cook, M., Krumlauf, R., Wiedemann, L.M., Sower, S.A., Decatur, W.A., Hall, J.A., Amemiya, C.T., Saha, N.R., Buckley, K.M., Rast, J.P., Das, S., Hirano, M., McCurley, N., Guo, P., Rohner, N., Tabin, C.J., Piccinelli, P., Elgar, G., Ruffier, M., Aken, B.L., Searle, S.M., Muffato, M., Pignatelli, M., Herrero, J., Jones, M., Brown, C.T., Chung-Davidson, Y.W., Nanlohy, K.G., Libants, S.V., Yeh, C.Y., McCauley, D.W., Langeland, J.A., Pancer, Z., Fritzsch, B., de Jong, P.J., Zhu, B., Fulton, L.L., Theising, B., Flicek, P., Bronner, M.E., Warren, W.C., Clifton, S.W., Wilson, R.K., Li, W., 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet 45, 415-421, 421e411-412.

Southard-Smith, E.M., Kos, L., Pavan, W.J., 1998. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. Nat Genet 18, 60-64.

Spokony, R.F., Aoki, Y., Saint-Germain, N., Magner-Fink, E., Saint-Jeannet, J.P., 2002. The transcription factor Sox9 is required for cranial neural crest development in Xenopus. Development 129, 421-432.

Stone, J.R., Hall, B.K., 2004. Latent homologues for the neural crest as an evolutionary novelty. Evol Dev 6, 123-129.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28, 2731-2739.

Taylor, K.M., Labonne, C., 2005. SoxE factors function equivalently during neural crest and inner ear development and their activity is regulated by SUMOylation. Dev Cell 9, 593-603.

van de Wetering, M., Oosterwegel, M., van Norren, K., Clevers, H., 1993. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. The EMBO journal 12, 3847-3854.

Van Otterloo, E., Li, W., Garnett, A., Cattell, M., Medeiros, D.M., Cornell, R.A., 2012. Novel Tfap2-mediated control of soxE expression facilitated the evolutionary emergence of the neural crest. Development 139, 720-730.

Wada, H., Makabe, K., 2006. Genome duplications of early vertebrates as a possible chronicle of the evolutionary history of the neural crest. Int J Biol Sci 2, 133-141.

Wagner, G.P., Amemiya, C., Ruddle, F., 2003. Hox cluster duplications and the opportunity for evolutionary novelties. Proceedings of the National Academy of Sciences of the United States of America 100, 14603-14606.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., et al., 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 79, 1111-1120.

Watanabe, K., Takeda, K., Yasumoto, K., Udono, T., Saito, H., Ikeda, K., Takasaka, T., Takahashi, K., Kobayashi, T., Tachibana, M., Shibahara, S., 2002. Identification of a distal enhancer for the melanocyte-specific promoter of the MITF gene. Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society 15, 201-211.

Wegner, M., 1999. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res 27, 1409-1420.

Wegner, M., Stolt, C.C., 2005. From stem cells to neurons and glia: a Soxist's view of neural development. Trends in Neurosciences 28, 583-588.

Westerfield, M., 2007. THE ZEBRAFISH BOOK, 5th Edition; A guide for the laboratory use of zebrafish (Danio rerio). University of Oregon Press, Eugene.

Yan, Y.L., Miller, C.T., Nissen, R.M., Singer, A., Liu, D., Kirn, A., Draper, B., Willoughby, J., Morcos, P.A., Amsterdam, A., Chung, B.C., Westerfield, M., Haffter, P., Hopkins, N., Kimmel, C., Postlethwait, J.H., 2002. A zebrafish sox9 gene required for cartilage morphogenesis. Development 129, 5065-5079.

Yan, Y.L., Willoughby, J., Liu, D., Crump, J.G., Wilson, C., Miller, C.T., Singer, A., Kimmel, C., Westerfield, M., Postlethwait, J.H., 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132, 1069-1083.

Yu, J.K., Meulemans, D., McKeown, S.J., Bronner-Fraser, M., 2008. Insights from the amphioxus genome on the origin of vertebrate neural crest. Genome research 18, 1127-1132.

Zhang, G., Miyamoto, M.M., Cohn, M.J., 2006. Lamprey type II collagen and Sox9 reveal an ancient origin of the vertebrate collagenous skeleton. Proceedings of the National Academy of Sciences of the United States of America 103, 3180-3185.

Zhang, J.Z., 2003. Evolution by gene duplication: an update. Trends in Ecology & Evolution 18, 292-298.

Figure 1

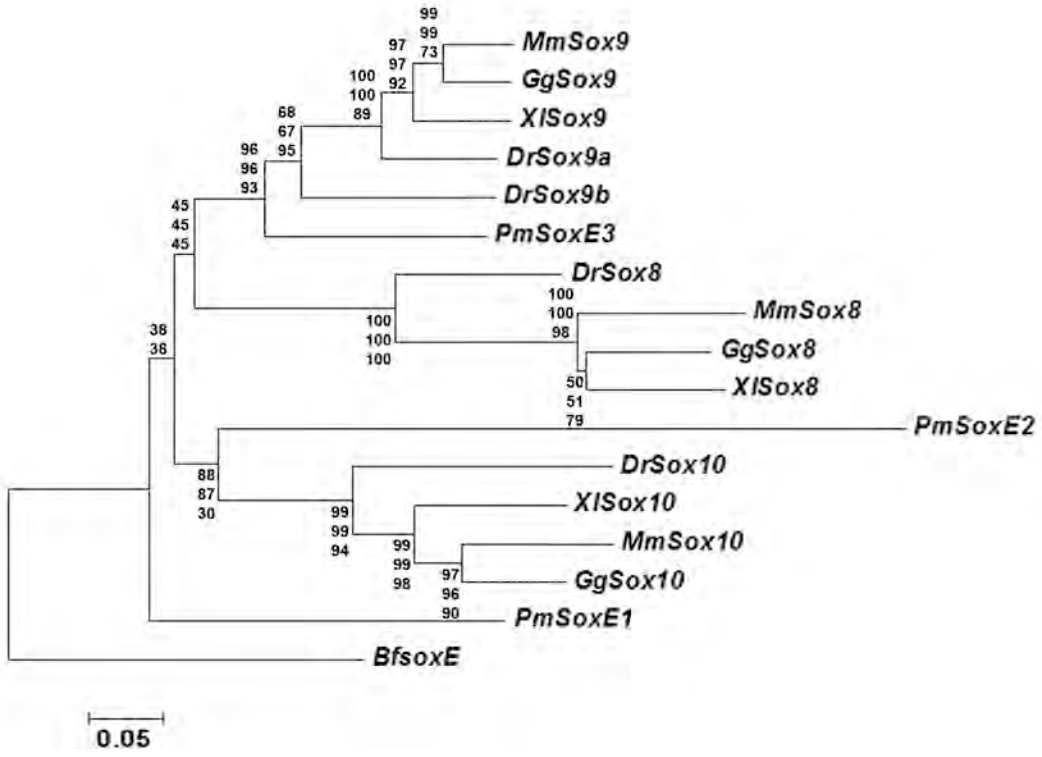

MmSox9
GgSox9
XlSox9
DrSox9a
DrSox9b
PmSoxE3
DrSox8
MmSox8
GgSox8
XlSox8
PmSoxE2
DrSox10
XlSox10
MmSox10
GgSox10
PmSoxE1
BfsoxE
0.05

Figure 2

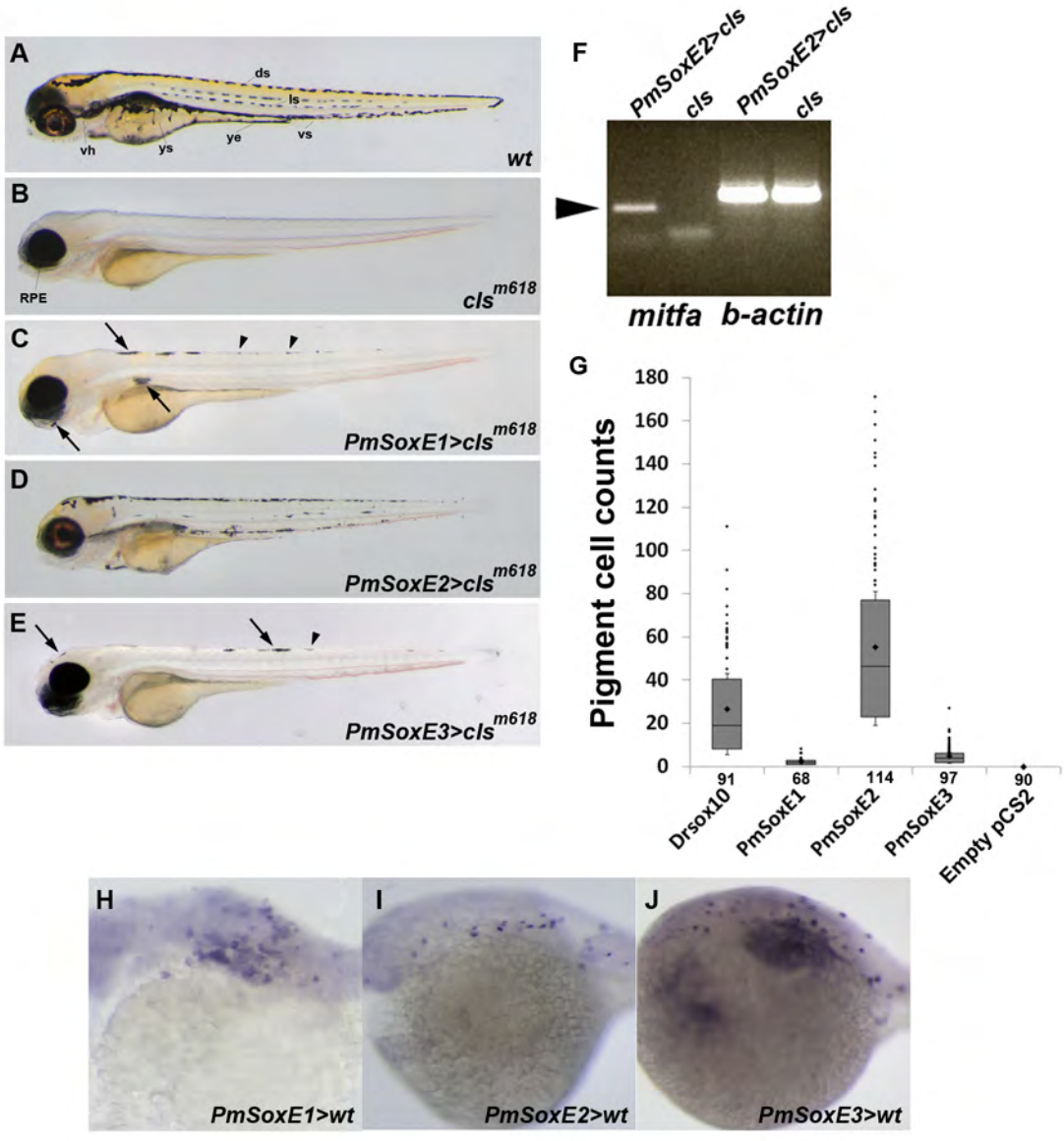

A
ds
ls
vh
ys
ye
vs
wt
B
RPE
cls m618
C
PmSoxE1>cls m618
D
PmSoxE2>cls m618
E
PmSoxE3>cls m618
F
PmSoxE2>cls
cls
PmSoxE2>cls
cls
mitfa b-actin
G
Pigment cell counts
180
160
140
120
100
80
60
40
20
0
91
68
114
97
90
Drsox10
PmSoxE1
PmSoxE2
PmSoxE3
Empty pCS2
H
PmSoxE1>wt
I
PmSoxE2>wt
J
PmSoxE3>wt

Figure 3

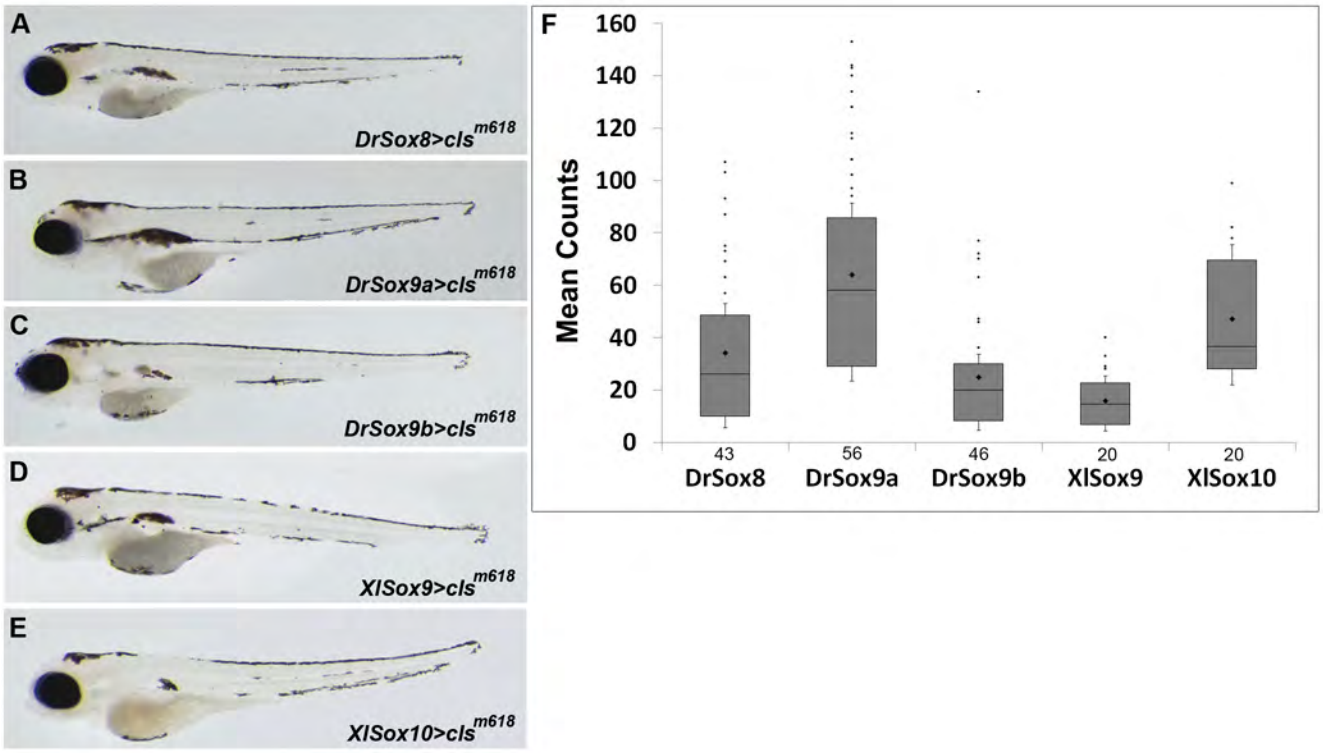

A
DrSox8>cls^m618
B
DrSox9a>cls^m618
C
DrSox9b>cls^m618
D
XlSox9>cls^m618
E
XlSox10>cls^m618
F
Mean Counts
160
140
120
100
80
60
40
20
0
43
56
46
20
20
DrSox8
DrSox9a
DrSox9b
XlSox9
XlSox10

Figure 4

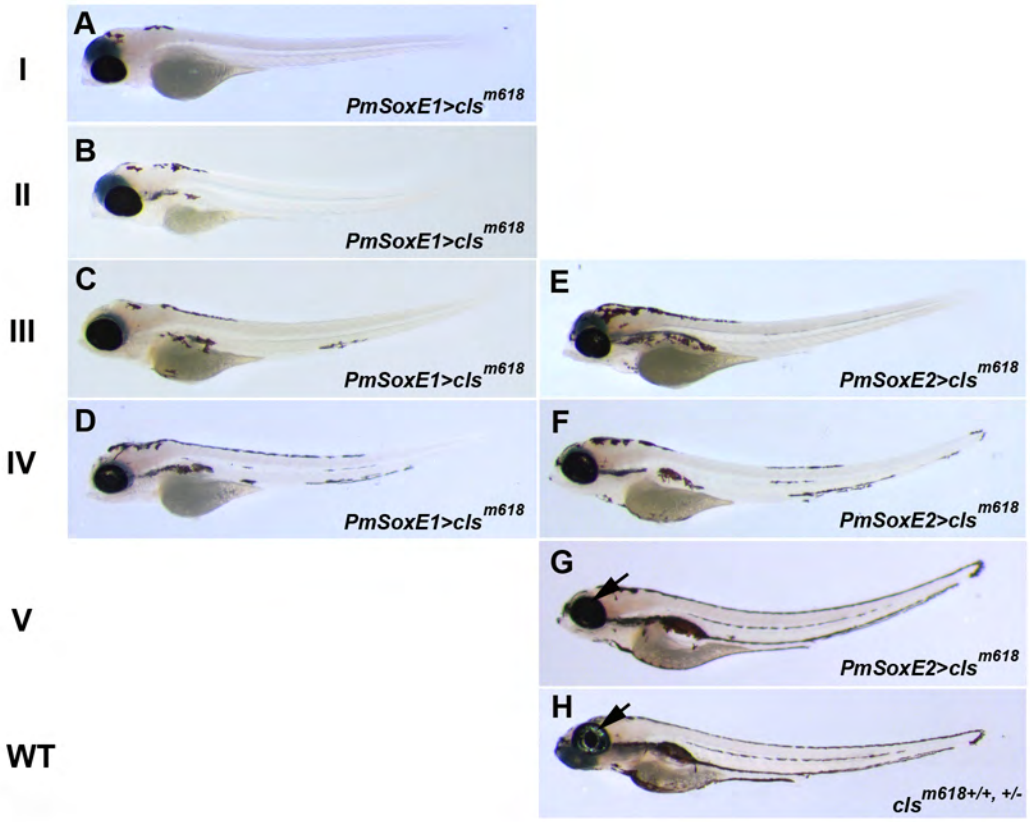

I
II
III
IV
V
WT
A
PmSoxE1>cls m618
B
PmSoxE1>cls m618
C
PmSoxE1>cls m618
D
PmSoxE1>cls m618
E
PmSoxE2>cls m618
F
PmSoxE2>cls m618
G
PmSoxE2>cls m618
H
cls m618+/+, +/-

Figure 5

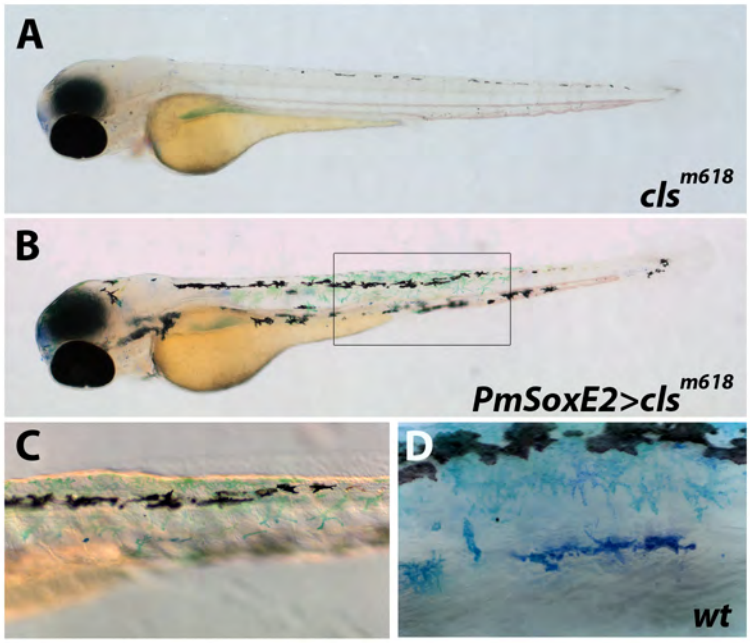


Figure 6

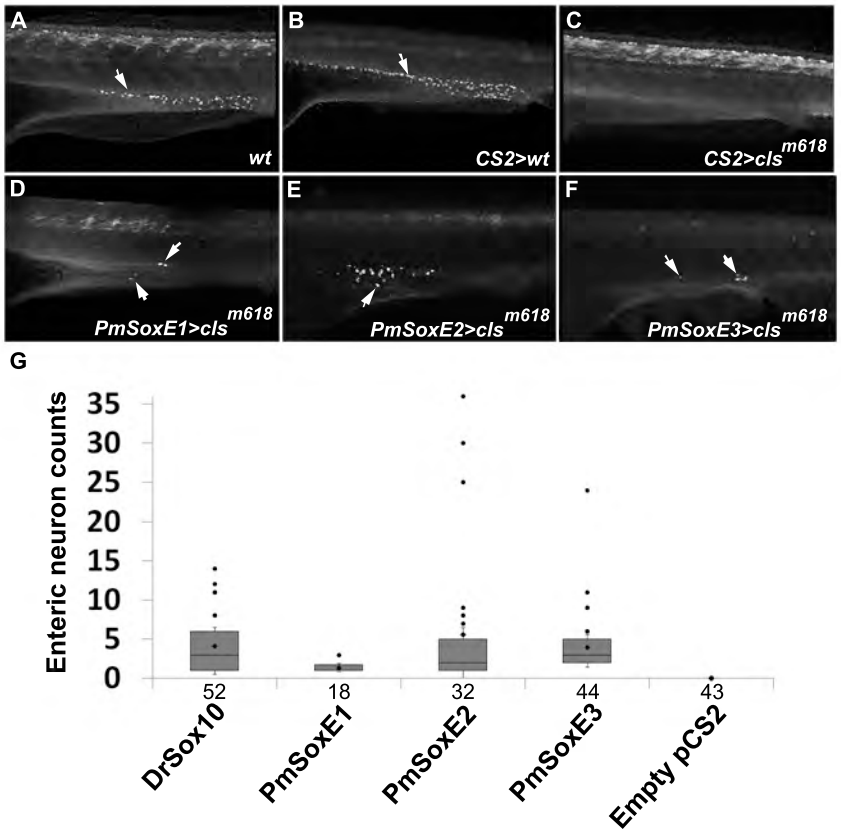

A
wt
B
CS2>wt
C
CS2>cls m618
D
PmSoxE1>cls m618
E
PmSoxE2>cls m618
F
PmSoxE3>cls m618
G
Enteric neuron counts
35
30
25
20
15
10
5
0
52
18
32
44
43
DrSox10
PmSoxE1
PmSoxE2
PmSoxE3
Empty pCS2

Figure 7

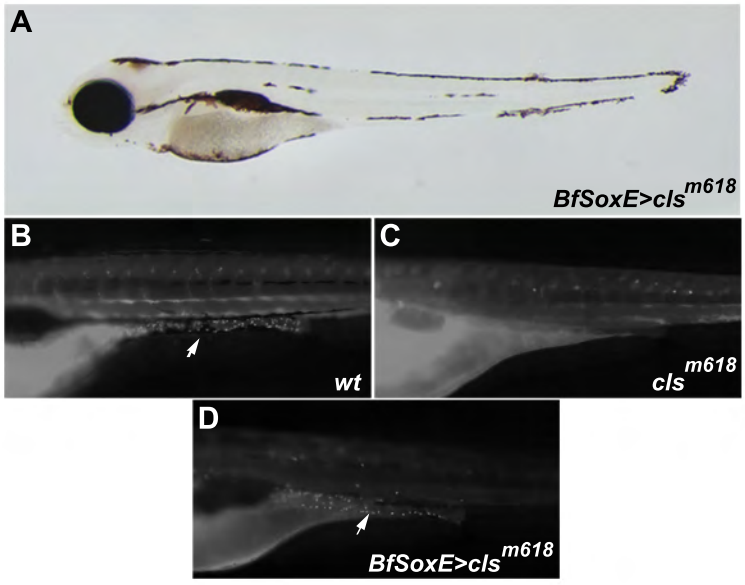


Figure 8

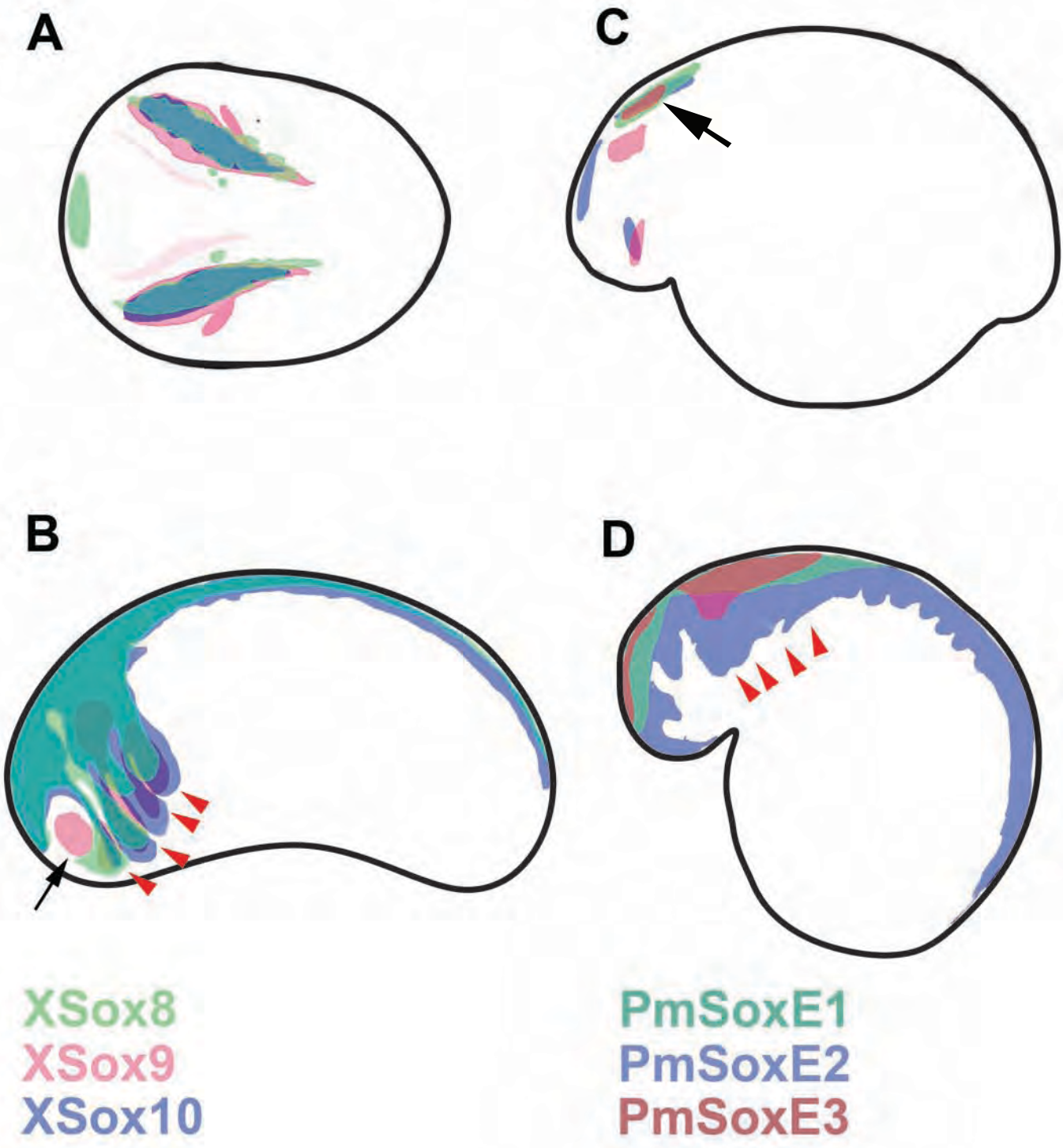


Table 1. pCS2-SoxE melanophore rescue in *cls* mutants

	# of experimental mutants observed	# of mutants with pigment	% rescue
Drsox10	98	92	94%
PmSoxE1	127	32	25%
PmSoxE2	107	99	93%
PmSoxE3	181	54	30%

Table 2. Tol2-SoxE melanophore rescue in *cls* mutants

	PmSoxE1	*PmSoxE2*
I	20	0
II	18	0
III	21	5
IV	10	19
V	0	51
No Pigment	44	3
# m618	113	78*
# injected	454	394

* the number of PmSoxE2-rescued m618 mutants, determined by the presence of iridiphores, is less than the predicted number based on an expected 1:3 Mendelian ratio (99/374)

Table 3. pCS2-SoxE ENS rescue in *cls* mutants

	# of experimental mutants observed	# of mutants with ENS	% rescue
Drsox10	108	52	48%
PmSoxE1	102	10	10%
PmSoxE2	83	24	29%
PmSoxE3	88	41	47%

Supplementary material Figure S1. (A) Phylogenetic relationships among chordates used in this study. The non-vertebrate chordate, Amphioxus, possesses a single *SoxE* gene. Lampreys are positioned basal to all jawed vertebrates and possess three SoxE genes, indicating SoxE gene duplication occurred early in vertebrate evolution. (B) Vertebrate SoxE functions mapped onto the SoxE phylogeny shown in Figure 1. Functions of *SoxE* proteins that arose prior to gene duplication are indicated by Roman numerals; basal vertebrate functions are indicated in parentheses; SoxE functions specific to agnathans or gnathostomes are shown in brackets. Adapted from Lakiza et al. 2011.

Supplementary material Figure S2. Alignment of lamprey *SoxE1*, *SoxE2*, and *SoxE3* to gnathostome *Sox8*, *Sox9*, and *Sox10* genes. Clustal X was used to create a full length alignment that was manually edited using BioEdit. Accession numbers are provided in materials and methods. Aligned conserved domains are shown in green (dimerization), blue (HMG), red (K2) and yellow (transactivation) boxes.

Supplementary material Figure S3. Lamprey *SoxE2* expression in wildtype zebrafish causes excessive pigmentation. (A) pigmentation in a 56 hpf zebrafish embryo. (B) Excess pigmentation, especially in the trunk region, in a 56 hpf zebrafish embryo following injection with *SoxE2.* Orientation: anterior facing left.

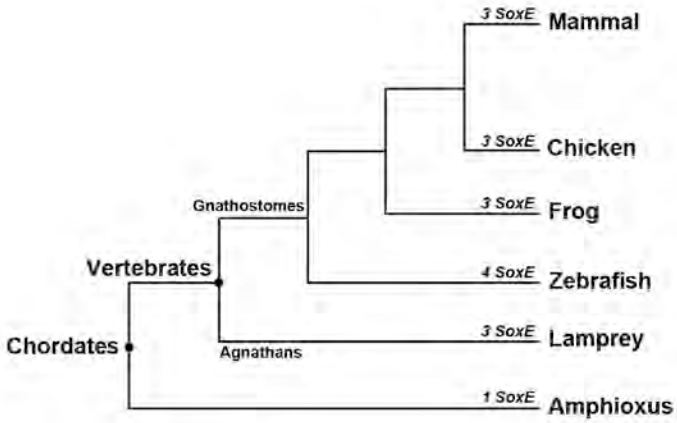
Supplementary material Figure S4. Amphioxus *SoxE* mRNA injection induces expansion of *Slug* in *Xenopus*. (A – B) *in situ* hybridization in *Xenopus laevis* embryos (stage 17) indicates expansion of *Slug* expression unilaterally on the side of the embryo injected with *BfSoxE* mRNA (arrow).

Supplementary material Figure S5. GFP expression in wildtype zebrafish embryos injected with either *neGFP* mRNA or CS2:*neGFP* plasmid. (A – C) 24 – 28hpf embryos show similar mosaic expression irrespective of mRNA or plasmid injection, highlighting the effectiveness of plasmid injection.

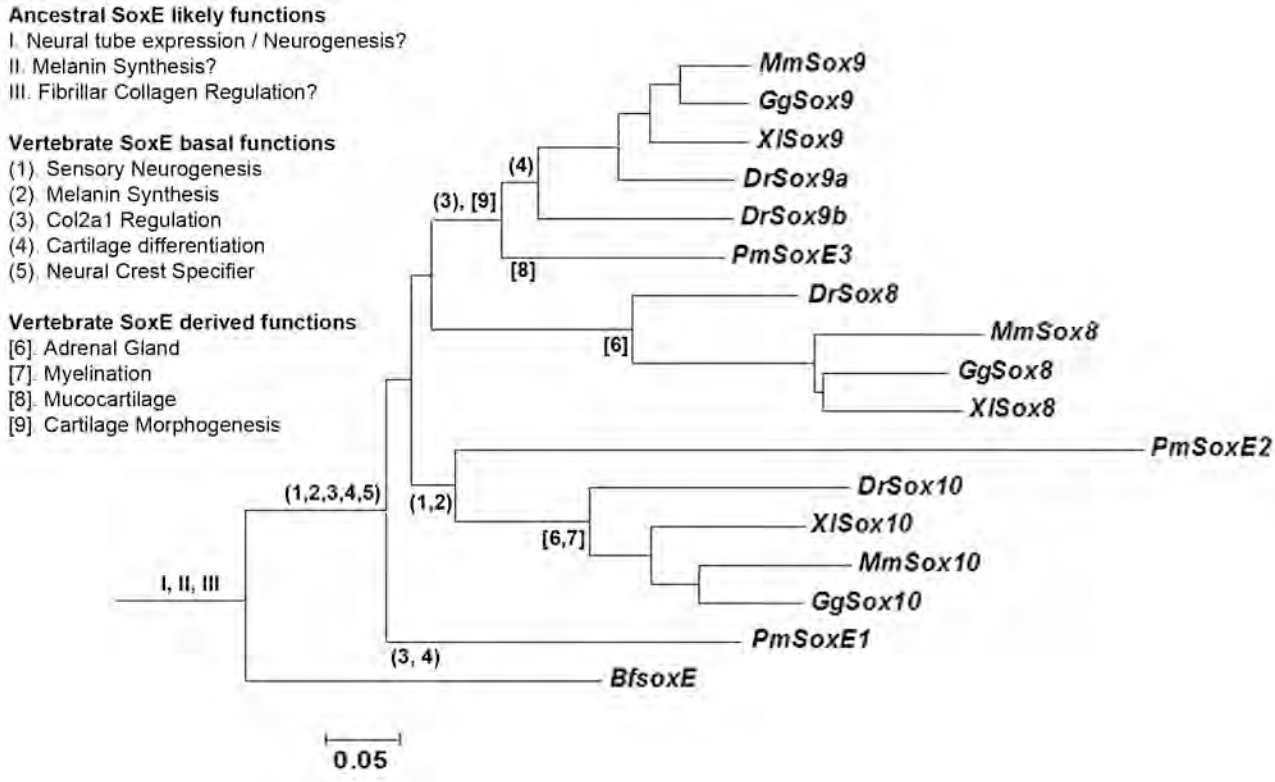
Supplementary material Figure S6. Melanogenesis following injection of PmSoxE2-Myc fusion constructs. PmSoxE2 was cloned into CS2+ vectors with 6 myc epitope repeats located at either the 3' or 5' end of the open reading frame. A comparison of pigment cell rescue among PmSoxE2, PmSoxE2-3'myc, and PmSoxE2-5'myc indicates that presence of the myc epitope in either orientation interfered with melanogenic activity of PmSoxE2.

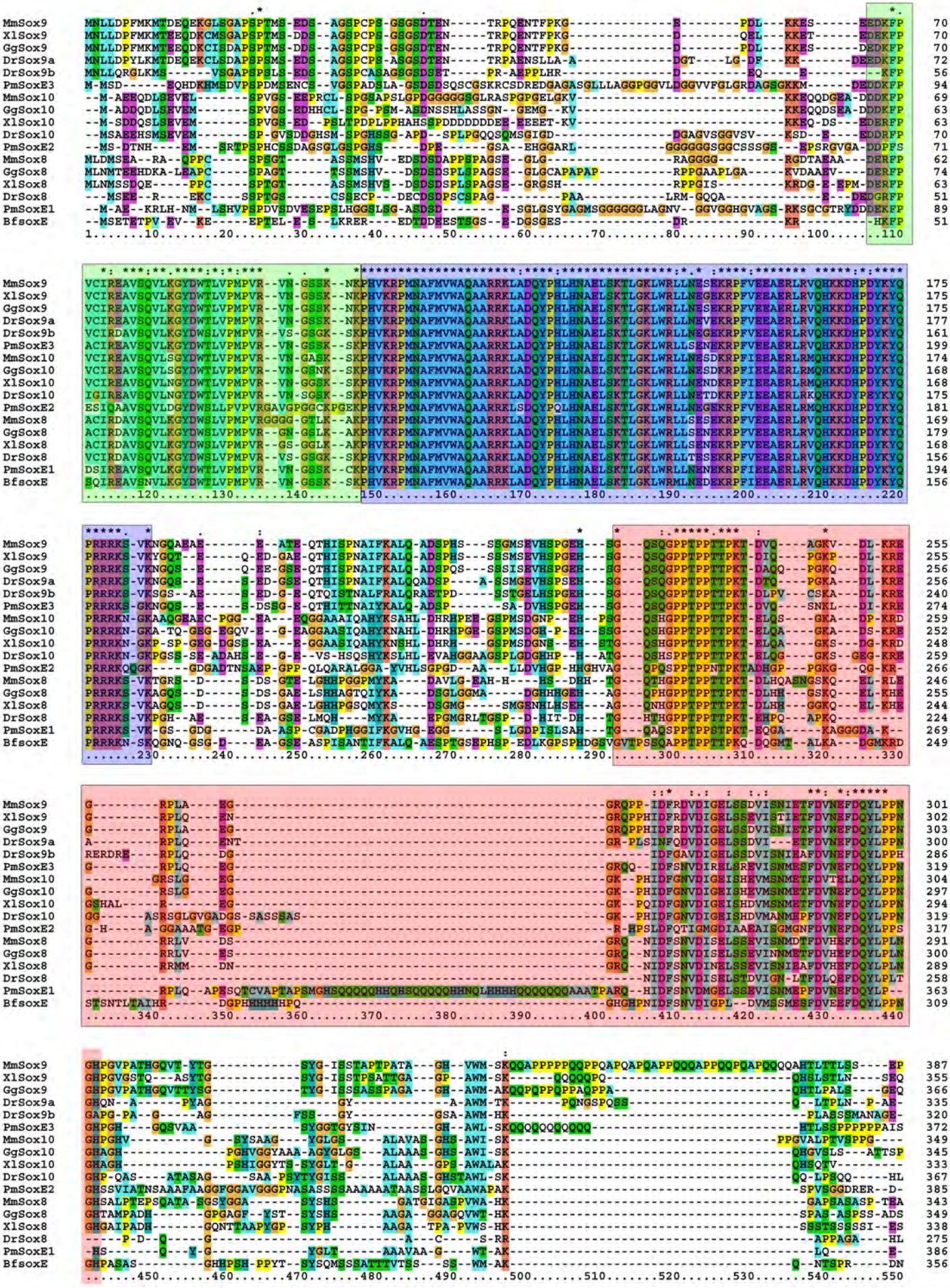
Supplementary Figure 1

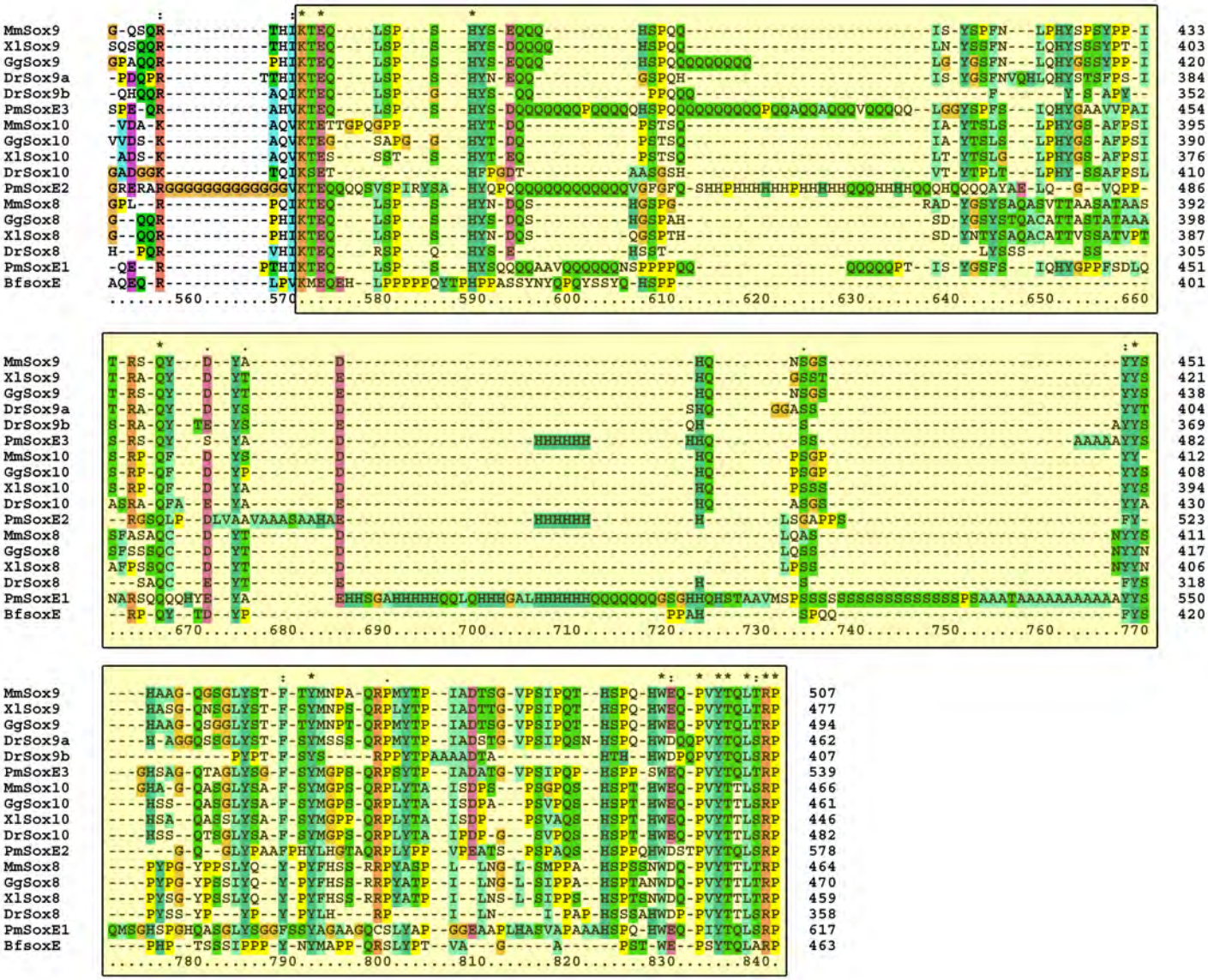
A

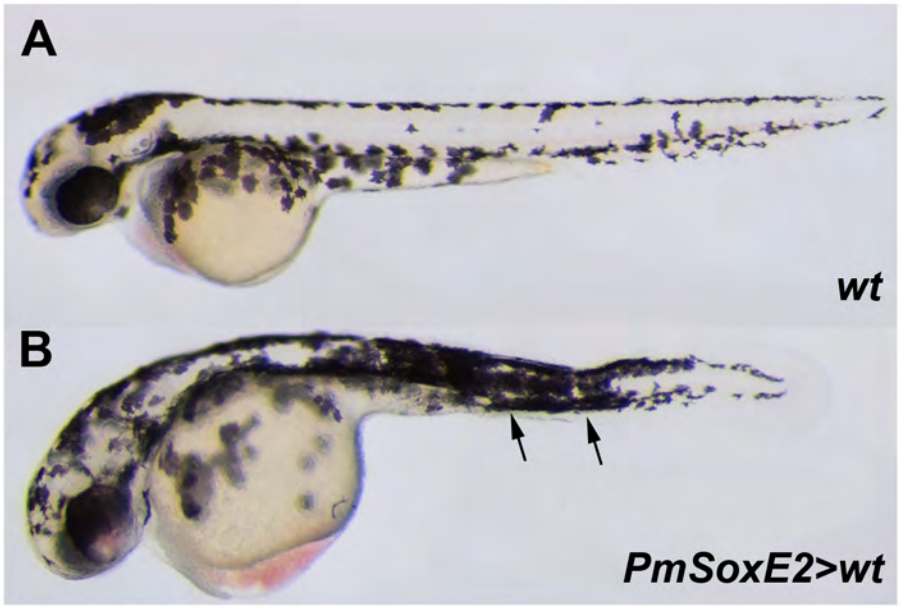


B

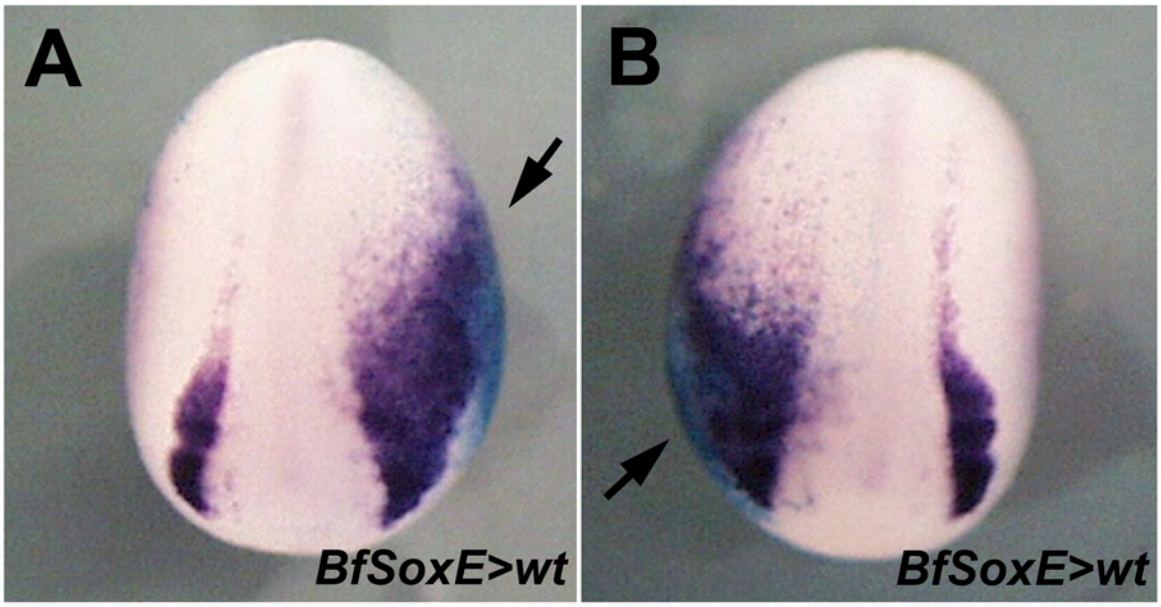




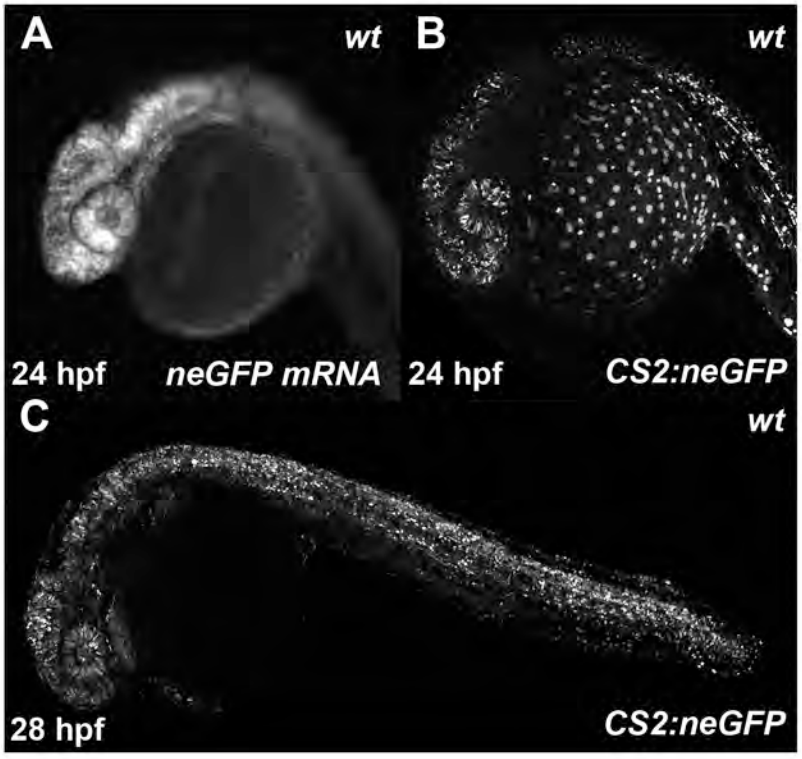

MmSox9
XlSox9
GgSox9
DrSox9a
DrSox9b
PmSoxE3
MmSox10
GgSox10
XlSox10
DrSox10
PmSoxE2
MmSox8
GgSox8
XlSox8
DrSox8
PmSoxE1
BfsoxE


A
wt
B
PmSoxE2>wt

Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6